DOCUMENT RESUME

ED 212 661 TM 820 057

AUTHOR	Quellmalz, Edys; Capell, Frank
TITLE	Defining Writing Domains: Effects of Discourse and Response Mode.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 79
GRANT	OB-NIE-G-78-0213
NOTE	134p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Correlation; Essay Tests; *Literary Genres; Models; Multiple Choice Tests; *Scores; Secondary Education; *Test Format; Testing Problems; Test Validity; Writing (Composition); *Writing Evaluation

ABSTRACT

The purpose of this study was to examine the stability of measures of student writing performance across types of discourse (genres) and across response modes (selected response: multiple choice; constructed response: single paragraph, and full length essay). The study addressed the following: (1) the relationship/stability of writing scores within and between genres when students are tested on different occasions (on writing tasks in the same discourse mode, in different discourse modes); and (2) the relationship/stability of writing scores across selected response items (multiple choice items), brief constructed response items (single paragraph task), and longer constructed response item (full essay 1-2 pages, 3-5 paragraphs). Results indicated that the relationship between a student's scores on essays written in the same mode of discourse is stronger than the correlation between two essays in different discourse modes. These findings suggest that generalizations about student writing competency should take care to reference the discourse domain rather than the more general domain of writing, as students' command of different discourse structures vary. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212661

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE - November 1979

STUDIES IN MEASUREMENT AND METHODOLOGY

Work Unit 1: Design and Use of Tests

Project Directors:
Eva L. Baker
Edys Quellmalz

Defining Writing Domains: Effects of Discourse and Response Mode

Edys Quellmalz and Frank Capell

(This work is submitted in fulfillment of the Final Report - Writing Study #5 as listed in grant)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. S. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant Number
OB-NIE-G-78-0213

TM 820 057

The work upon which this publication is based was performed pursuant to a grant with the National Institute of Education, Department of Health, Education and Welfare. Points of view or opinions stated do not necessarily represent official NIE position or policy.

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

Table of Contents

Introduction 1

Method 4

  Sample. 4

  Treatments. 4

  Procedures. 7

  Measures. 10

Findings 21

Discussion 60

References 65

Appendices A-E

Activity 5

Final Report

DEFINING WRITING DOMAINS
EFFECTS OF DISCOURSE AND RESPONSE MODE

Factors affecting large scale assessment of student writing ability are numerous and complex. Conventionally, assessment programs aimed at district, state or national populations have measured writing skills with objective tests. Support for such indirect measures derives from the high correlations between objective tests and writing samples, the erratic reliabilities accompanying impressionistically scored essays and economic and logistical demands of collecting and scoring writing samples (Coffman, 1971; Godshalk, 1966; Breland, 1977). More recently demands for writing tests with content, construct and ecological validity have prodded reinstatement of written production tasks. Too often, however, large scale assessments collect only one writing sample, despite the well documented fluctuation of writing performance from one sample to the next (Braddock, 1963; French, 1962; Diederich, 1974). By necessity, a single sample taps student performance on only one type of discourse and on one topic. This limitation is a problem since the elements of particular forms of discourse differ from one another. Salient structural features of argument, for instance, are issues, reasons, and conclusions, while stories include plot, character, setting, and theme. Exposition involves main idea, supporting detail and logical development; narrative presents details in chronological order, and description portrays concrete details in spatial order. As the defining structural elements of the modes of discourse differ (Kinneavy, 1971) so it seems likely that the schemata or frames activated in the writer by a writing task should differ (Anderson, 1977; Minsky, 1976). Statements of writing test results generalizing about student competency on a text element such as organization, then, should be limited to the particular form(s) of that element characteristic of that discourse mode. Students may well be more skilled at narrative writing tasks requiring chronological organization, than descriptive or expository tasks requiring spatial or logical organization. Thus the profile of writing ability for a student derived from a writing test calling for exposition may differ from the profile of writing ability for the same student derived from a descriptive, narrative or persuasive essay. The potential influence of mode of discourse on a writer's performance has been discussed by Braddock, Lloyd-Jones and Schoer (1963). Veal and Tillman (1971) examined discourse effects in elementary children's writing and found the expository mode to differentiate most among young writer's skills. More recently the Primary Trait scoring systems developed by Lloyd-Jones are referenced to the structural properties of different genres (Lloyd-Jones, 1977).

Along with the issue of skill comparability across types of writing, or genres, is the question of the response form or measurement mode in which writing skill should be assessed. While many claims are made for the predictive or concurrent validity of objective writing measures, (Coffman, 1971; Breland, 1977) indirect measures simply are not considered by writing researchers to meet the more crucial standards of content or construct validity (Braddock, et al., 1963; Cooper and Odell, 1977). Multiple choice tests provide some valuable information, but are, after all, measures of reading comprehension, not measures of actual writing ability. As such, recognition tasks are often considered, at best, enroute performances to constructed responses (Bourne, 1966).

The accuracy of any single essay as a reflection of writing ability is certainly not without peril, and essays are time consuming and expensive to score (Kincaid, 1963; Diederich, 1974). A question also arises, then, about the information yielded by shorter writing samples. To what extent will a single paragraph provide acceptable information about writing ability comparable to essay performance? A comparison of short and long written samples and multiple choice scores could furnish valuable information about the utility of different methods of evaluating writing skill.

The purpose of this study was to examine the stability of measures of student writing performance across types of discourse (genres) and across response modes (selected response: multiple choice; constructed response: single paragraph, and full length essay).

The study addressed the following research questions:

(1) What is the relationship/stability of writing scores within and between genres when students are tested on different occasions
   (a) on writing tasks in the same discourse mode?
   (b) on writing tasks in different discourse modes?

(2) What is the relationship/stability of writing scores across:
   -selected response items (multiple choice items)?
   -brief constructed response items (single paragraph task)?
   -longer constructed response item (full essay 1-2 pages, 3-5 paragraphs)?

METHOD

To examine the relationship of writing scores yielded by differing discourse and response mode requirements, high school students received writing tests on three separate occasions. Each student received a multiple choice test and a paragraph writing task as well as two full length essay assignments. Ratings of the essays and paragraph on an analytic scale and scores on the objective test provided the basis for the comparisons of interest.

Sample

Participants were students from all three high schools in a small high school district in the Los Angeles area. Socioeconomic status in the district ranged from upper-lower to upper-middle class. Teachers of four English or composition classes within each school volunteered to participate in the study. Because of the research questions comparing paragraph and essay production, average and above average classes were requested. Students were eleventh or twelfth graders, judged by their teachers to be in average to above average classes. Table 1 indicates the composition of classes included in the study.

(Insert Table 1 here)

Means and standard deviations of student scores on related verbal scores appears in Table 2.

(Insert Table 2 about here)

In all cases the mean scores for the sample fall within "average" ranges for each test.

Treatment Groups

Students within each class were randomly assigned to one of four major testing conditions defined by the discourse mode(s) of the essay tasks. Within the Same Genre condition students received *either* two expository essay assignments on topics 1 & 2 and an expository paragraph *or* two nar-

Table 1

Sampling Design

School	Classroom	Number of students	Class tracking level
1	1	27	1
	2	33	2
	3	37	3
	4	33	3
2	1	29	3
	2	25	1
	3	28	1
	4	26	2
3	1	20	1
	2	30	3
	3	26	2
	4	26	3
Total 3	12	340	

1 = Advanced, Honors
2 = Above Average
3 = Average

TABLE 2

Means and Standard Deviations of Students

	DAT-V	SFTAA	PreSAT	SAT	Spring English	Fall English
	%	%				
$\overline{X}$	61	72	445	516	2.85	2.78
SD	29	24	109	119	.82	.78
n =	225	82	161	80	201	205

DAT-V = Differential Aptitude Test - Verbal

SFTAA = Short Form Test of Academic Aptitude

PreSAT = Preliminary Scholastic Aptitude Test

SAT = Scholastic Aptitude Test

Spring English = Spring 1977 English grades

Fall English = Fall 1978 English grades

rative essay assignments on topics 3 & 4 and a narrative paragraph. Within the Different Genre condition students wrote a narrative and an expository essay, on either topics 1 & 4 or 2 & 3. Half of these subjects wrote the expository paragraph, while half wrote the narrative paragraph. The within-subject variable, response mode, consisted of the multiple choice test (selected response mode), the paragraph (short constructed response mode) and the essay (long constructed response). During the three testing occasions subjects received the multiple choice test and paragraph on one occasion, and an essay on each of the other two occasions. The tasks were administered in two orders: Order 1: multiple choice plus paragraph, essay 1, essay 2; or Order 2: essay 1, essay 2, multiple choice test plus paragraph. Also, the order of the paragraph and multiple choice test was counterbalanced, as was the order of the two essays. Table 3 presents the treatment conditions employed in each classroom.

(Insert Table 3 here)

Procedures

Data collection for all twelve classes occurred over a two week period near the beginning of the Spring semester. CSE research personnel administered all tests.

Three separate packets of materials, one packet for each measurement occasion, were assembled for each class. The packets contained a class roster with students' names, code number and treatment group; a script for the data collector; directions concerning administration of the measures; copies of the measures; and extra pens. To ensure that students received the appropriate measure according to treatment group assignment, each copy

Table 3

Study Design

	Same Genre				Different Genre			
Essay:	ED, EV		ND, NV		ND, EV		NV, ED	
(Order:)	1	2	1	2	1	2	1	2
Paragraph:								
Expository			X	X	x	x	x	x
Narrative	X	X			x	x	x	x
Multiple-Choice:								
Expository								
Narrative								

ED = Expository essay on drugs

EV = Expository essay on disrespect/violence

ND = Narrative essay on drugs

NV = Narrative essay on disrespect/violence

Order 1 = Paragraph and m/c Test before essays

Order 2 = Paragraph and m/c Test after essays

X = empty cell

x = ½ subjects in cell not
receiving paragraph in
that genre

was labeled with the student's name, code number, treatment group, and the date on which the measure was to be administered. The objective tests were printed on white paper, the essay directions on colored paper to facilitate the ease with which directions were given. To allow for subsequent photocopying, lined paper cut to 7" x 10" was provided for writing the essays.

Prior to the first data collection activity, teachers were asked whether they wished to have copies of their students' writing samples and, if so, whether they intended to evaluate the samples as part of the students' course grade. All teachers wished to receive copies of the samples, but none intended to use them for grading purposes. Xerox copies of the collected samples along with ratings and the multiple choice scores were returned to the teachers within a month. Students in all classes were informed that their essays would be read and reviewed by their English teachers.

Procedures for all classes for each of the three measurement occasions were similar, with the exception of Occasion One when the data collector gave a brief introduction to the study. Students were assured that no student, teacher, or school names would be reported. The data collector asked that the class remain quiet until the end of the period and explained that questions would be answered individually. After the measures were distributed the data collector first gave directions for writing the essays, followed by directions for completing the objective measure and paragraph. Students then were requested to begin working.

The remainder of the class period was devoted to completing the measurement tasks, with an announcement of the time by the data collector when ten minutes were left. The amount of time allowed ranged from thirty-five to forty-five minutes according to variations in the schools' schedules. The average amount of time was approximately forty minutes.

Measures

Writing Samples. Specifications controlling the essay and paragraph assignments delineated the essay genre, topic attributes and particular directions to students cueing desired essay elements. The definitions for exposition described by Spooner-Smith (1978) were employed in the expository essay domain specifications. Directions were devised to cue desired essay components including main idea, organization and support. The specifications were then used to generate the expository essay tasks. The "drug" topic used by Spooner-Smith (1978) and Winters (1978) was adapted by eliminating reference to alcohol and smoking. A second topic, juvenile disrespect for adults, was also developed. Domain specifications for the expository essay tasks appear in Appendix A along with the two essay topics.

The narrative domain definition drew upon, but differed from that used by Pitts (1978). Rather than imaginative narration/description, the present study attempted to elicit "factual functional narrative" writing where the literary components of character, setting, and conflict were deferred in favor of straightforward accounts of what happened over a period of time. In an attempt to provide a basis for examining the relationship of discourse - specific writing skills to topic knowledge, the same two topics used for the expository essay were used for the narrative essay. Directions specified a narrative writing purpose (tell what happened over a period of time) and the desired essay components of main idea, organization and support.

Appendix B presents domain specifications for the narrative essay tasks and the two narrative topics.

A paragraph topic, alcohol, was couched in either expository or narrative directions, and instructions specified that only a paragraph was to be written. Paragraph task specifications were the same as those for the essay with the exception that the directions asked students to write a paragraph. The paragraph topics appear in Appendix C.

All writing topics and directions were field tested with three high school classes and retained for the final study. Topics were retained if pilot essays seemed to indicate that students were knowledgeable about the topic and could write in the genre requested.

Rating Scales. Two analytic rating scales were developed. The Expository Scale, used by Spooner-Smith (1978), was adapted for the questions of interest in the present research. As a result of the Spooner-Smith (1978) and Winters (1978) studies, the new expository scale added a subscale for General Impression, in which raters were directed to read a paper first quickly and rate it according to their global judgment of its quality as an example of exposition. The Spooner-Smith subscales of Paragraph Development and Paragraphing were incorporated into a single Organization subscale. The Expository Scale in this study, then, had five elements: General Impression, Focus/Main Idea, Organization, Support, and Mechanics.

For the Factual Narrative analytic scale, five parallel subscales were developed. Referencing the scale developed for Imaginative narrative/descriptive writing, the new scale included subscales for General Impression, Focus, Organization, Support (incorporating some reference to character description, background and setting descriptions) and Mechanics. These scales and their abbreviated versions are in Appendix D.

Multiple-Choice Test. Stimulus attributes delineated in the writing sample specifications were used to develop the passages read in the multiple-choice test. Ten passages of 4-6 sentences, approximately 200 words, were generated drawing from some of the items used by Spooner-Smith in 1979. Five passages were expository; five were narrative. In addition to the criteria used by Spooner-Smith, the expository passages varied according to the development structure employed. For each passage there were three questions related to main idea, organization, and support. Main idea questions were referenced to a stated generalization near the beginning or end of the passage. Organization questions required selection of a new sentence which would fit best at a point in the passage marked by an arrow. Support questions asked which new sentence would also best support the passage main idea. The test, then, consisted of 30 items, fifteen related to exposition, fifteen related to narration. Five questions relating to expository main idea, and five relating to narrative main idea allowed comparisons by structural element (main idea, organization, support) within and between genre conditions. Specifications for the test and a copy of the test are in Appendix E. Table 4 describes the design of the multiple choice test.

The objective test was also field tested and revised for use in the final study.

(Insert Table 4 here)

Rating Procedures

Students completed three writing samples, two essays and a paragraph.

Table 4

Multiple Choice Test Design

Genre	Item Numbers	Number of Items
Exposition	1 - 15	15
Main Idea	1, 4, 7, 10, 13	5
Organization	2, 5, 8, 11, 14	5
Support	3, 6, 9, 12, 15	5
Narration	16 - 30	15
Main Idea	16, 19, 22, 25, 28	5
Organization	17, 20, 23, 26, 29	5
Support	18, 21, 24, 27, 30	5
	Total	30

Students who did not complete both full essays were not included in the analysis. Photocopies of the writing samples were prepared for the raters.

The rating procedures involved training sessions in the use of the Expository and Factual Narrative Scales, pilot tests of rater agreement, then rating of essays and paragraphs. Four instructors from the UCLA English and Subject A Departments scored the writing samples. All raters had previous experience evaluating high school and freshman compositions of students ranging from low to high ability writers. Raters were paid for their participation. Following the rating procedures described by Spooner-Smith (1978) and Winters (1978), all four raters worked with the researcher on approximately 20 training papers for approximately 8 hours to use the subscales consistently. Appendix F reports the pilot test statistics for determining rater agreement on two expository essays by twenty students.

Both coefficient alpha and generalizability statistics were calculated. The alpha coefficients ranging from .86 to .92 and estimated variance components attributable to raters ranging from .03 to .09 indicated that the training had promoted acceptable levels of rater consistency.

The same procedure was then followed for the Factual Narrative Scale. Training time was approximately 6 hours. Appendix F also presents the pilot test statistics results for that scale. Ratings on the Mechanics subscale were divergent (alpha = .73), so the Mechanics scores of a set of the pilot papers were further discussed. An additional set of ten papers were then rated on the Mechanics subscale. No ratings differed by more than one point; therefore rating procedures began for the final study scoring.

The four raters were divided into two pairs. One pair began rating narrative essays, the other pair began rating expository essays. One half the students from all testing conditions were randomly assigned to rater pairs. Each pair read both essays written by a student, though the order in which the essays were read was randomly arranged for each rater.

After each set of approximately 25 papers, checks were made for continuing rater agreement. The rater pair read a common paper, not included in the analyses, compared scores, and resolved any disagreements.

After each rater pair read student essays in one genre, the next set of papers in the other genre was read. Paragraphs grouped into narrative or expository sets were then read. Again, the order of genre set and individual student paper was randomized. Rating time was approximately 3 minutes per essay, 2 minutes per paragraph. Appendix G presents the final reliability statistics for each rater pair on each expository and narrative subscale.

In accordance with findings of previous writing assessments, ratings were less consistent on the final study papers. On the Expository scale Raters 1 and 2 standardized item (S. I.) alphas ranged from .51 to .81; their generalizability coefficients(G. C.) from .22 to .68. They differed most on scores relating to the General Impression subscale (S. I. alpha = .51, G. C. = .22) and the Mechanics subscale (S. I. alpha = .69, G. C. = .57). The S. I. alphas of Raters 3 and 4 ranged from .50 to .85, their G. C. from .67 to .85. They differed on the Support subscale (S. I. alpha = .50, G. C. = .85).

These statistics indicate that, despite the initially high reliabilities of raters immediately after the training session, the reliability checks performed after approximately every twenty-fifth paper during the final study ratings were not sufficient to combat "rater drift" on some of the subscales.

Measurement Properties of the Analytic Scale

Essay scores were calculated by averaging the scores assigned by both raters to each essay for each subscale. The range of possible points for each subscale was 1-4, with a total possible essay score of 20. Means and standard deviations for the essay ratings in each testing condition are displayed in Table 5.

(Insert Table 5 here)

Subscale means ranged from .88 (22%) to 2.56 (64%). Discussion of the variabilities will be included in the Findings section of this report.

Measurement Properties of the Objective Multiple Choice Test

For the multiple choice test, means, standard deviations, and reliabilities were calculated. The range of subscale points was 1-5 within genre subscale, 1-30 for the total test. Means, standard deviations, and average percent correct on the subsections and total test are displayed in Table 6.

(Insert Table 6 here)

Subscale means ranged from 3.31 (66%) to 4.61 (92%). Table 7 presents the reliabilities of the test subsections and the total test. Alpha reliability of the total test was .81; subsection reliabilities ranged from .39 to .59.

Table 5

Means and Standard Deviations of Essay Ratings

Test Condition:		Same Genre				Different Genre			
		ED	EV	ND	NV	ND	EV	NV	ED
GI	$\bar{X}$ =	2.20	2.01	1.27	.88	1.43	1.89	.94	2.09
	SD =	.60	.70	1.44	1.09	1.89	.66	.88	.54
Focus	$\bar{X}$ =	2.45	2.3	2.21	2.21	2.28	2.19	2.26	2.33
	SD =	.69	.64	.68	.64	.59	.69	.51	.53
Organization	$\bar{X}$ =	2.16	1.98	1.88	1.62	1.88	1.95	1.52	2.08
	SD =	.72	.70	.98	.71	.85	.70	.56	.56
Support	$\bar{X}$ =	2.34	2.38	2.42	2.19	2.36	2.28	2.10	2.26
	SD =	.64	.54	.81	.66	.70	.65	.54	.53
Mechanics	$\bar{X}$ =	2.35	2.34	2.56	2.42	2.28	2.03	2.35	2.42
	SD =	.64	.68	.69	.67	.80	.81	.56	.54
Total	$\bar{X}$ =	11.50	10.90	10.35	9.32	10.23	10.31	9.18	11.19
	SD =	2.78	2.67	3.84	2.89	3.36	3.02	2.02	2.05
	N =	40	40	39	39	40	40	54	54

Table 6

Means and Standard Deviations of Multiple-Choice Test

	Expository	Narrative	Total
Focus $\overline{X}$	4.61 (92%)	4.50 (90%)	9.13 (91%)
s.d.	.77	.89	1.38
Organs $\overline{X}$	4.05 (81%)	3.31 (66%)	7.39 (73%)
s.d.	1.03	1.13	2.03
Support $\overline{X}$	4.56 (91%)	4.41 (88%)	8.97 (90%)
s.d.	.77	.98	1.51
Total $\overline{X}$	13.23 (88%)	12.23 (82%)	25.33 (84%)
s.d.	2.02	2.43	4.26

N = 241

Questions relating to the expository passages had an overall reliability of .69; those relating to narrative passages, a reliability of .73.

(Insert Table 7 here)

Table 7

Multiple Choice Test Coefficients of Reliability

Subsection	E_f	E_o	E_s	E_t	N_f	N_o	N_s	N_t	T_f	T_o	T_s	T_t
Reliability	.57	.41	.39	.69	.51	.56	.59	.73	.67	.66	.68	.82
N	225	225	225	225	225	225	225	225	225	225	225	225

E_f = Expository focus

E_o = Expository organization

E_s = Expository support

N_f = Narrative focus

N_o = Narrative organization

N_s = Narrative support

$T_f = E_f + N_f$

$T_o = E_o + N_o$

$T_s = E_s + N_s$

T_t = Total test

FINDINGS

Within and Between Genres. The first research question was "What is the relationship of writing scores when students are tested on writing tasks either within the same discourse mode or on two different discourse modes?" Data relevant to this question first compared scores students received on the two essays which were either both of the same genre (two expository essays, two narrative essays) or in different genres (one narrative essay, one expository essay).

Essay Scores. Table 5 presents the means and standard deviations of the essay ratings for each testing condition. On all five subscales and on total essay scores, narrative ratings were lower than expository ratings. This finding may be due to the differential curricular emphasis given to narrative and expository writing in the schools, to subjects' lack of knowledge at a personal experience level with information required to deal with the narrative topics, or to raters' tendency to score narratives more stringently.

At the descriptive level, the relationship between the two essay means of students in the Same Genre condition is stronger than that between the means of students' two essays in the Different Genre condition.

Table 8 displays the correlations between students' two essay scores on each of the analytic subscales. In general, correlations between essay scores of students in the Same Genre condition are higher than those of students in the Different Genre condition.

(Insert Table 8 here)

Table 8

Correlation Between Student's Two Essays

	ED, EV	ND, NV	ND, EV	NV, ED
GI	.56	.39	.33	-.08
F	.43	.68	.41	.37
O	.42	.42	.20	-.10
S	.27	.28	.07	.38
M	.58	.68	.63	.50
Total	.60	.55	.41	.22
N =	40	39	40	54

GI = General Impression

F = Focus

O = Organization

S = Support

M = Mechanics

If one examines individual subscales, looking across the conditions, those that seem to differentiate most between the different discourse modes are General Impression and Organization. This finding might be expected, since General Impression requires a judgment about the global quality of the essay as an example of exposition or narration. Therefore the constellation of essay factors influencing this judgment should be the most comprehensive for each discourse mode and thus the most discriminating. Structurally, exposition and narration differ dramatically in their characteristic use of logical or temporal organizations, respectively. On the Mechanics subscale, correlations across conditions are more comparable, reinforcing the notion that the constellation of syntactic, punctuation, spelling and usage skills are not as discourse-mode specific.

To test the statistical difference between scores of students in the same and different genre conditions, the correlations in Table 8 were transformed to standardized scores, average standardized scores were calculated for the same and for different conditions, and these average standardized scores for same and different conditions were contrasted. Table 9 presents the results of this analysis.

(Insert Table 9 here)

The comparison reveals that the relationship between a student's two essay scores on General Impression, Organization, or Total Score is stronger when students write in the same genre than when they write in different genre.

Paragraph Scores. Comparisons between genre on the paragraphs were conducted only at the group level, since an individual student wrote only one paragraph. Table 10 presents the results of these comparisons. The Analytic Scales for Narration and Exposition were used for rating the para-

Table 9

Comparison Between Same and Different Genre Testing Conditions

	Same Genre			Different Genre			
	ED,EV	ND,NV		ND,EV	NV,ED		
	1	2		3	4		$Z_{observed}$
	z	z	$z_{ave.12}$	z	z	$z_{ave.34}$	
GI	.633	.412	.5225	.343	-.08	.1315	2.516*
F	.460	.829	.6445	.436	.388	.412	1.496
O	.448	.448	.448	.203	-.100	.0515	2.551*
S	.277	.288	.2825	.070	.400	.235	.306
M	.662	.829	.7455	.741	.549	.645	.647
Total	.693	.618	.6555	.436	.224	.33	2.095*
N	40	39	79	40	54	94	

$^*p < .05$

GI = General Impression
F = Focus
O = Organization
S = Support
M = Mechanics
T = Total

graphs. Subscale scores ranged from 1-4, total scores from 1-20. Ratings of Narrative paragraphs were generally lower than ratings of Expository paragraphs. Ratings of Expository paragraphs differed significantly from Narrative paragraphs on the General Impression, Focus, Organization subscales and on the Total scores, suggesting that as a population these high school students were more competent writing about the topics in the expository mode. Consonant with essay ratings, Mechanics and Support were not as influenced by the different discourse tasks.

(Insert Table 10 here)

Multiple Choice Test. On the Multiple choice test, comparisons of interest were the scores individuals received on the narrative and expository sections of the exam. Table 6 presents the means and standard deviations of the test. On a reading comprehension test of items measuring writing-related skills, students were able to answer focus/main idea and support questions similarly well for both expository and narrative passages. On Organization questions, however, students had more difficulty in general (73% overall average) and particularly with narrative organization (66%). Table 11 displays the correlations between individuals' expository and narrative scores on the three multiple choice subscales. While correlations are statistically significant, the correlations between the Narrative and Expository subscales are substantially lower (.47, .49) than the correlations between the total Expository and Narrative scores (.65).

(Insert Table 11 here)

TABLE 10

T-Test, Means and Standard Deviations of Expository and Narrative Paragraphs

		Expository Paragraph	Narrative Paragraph
General Impression	$\overline{X}$	1.93	1.05
	S.D.	.71	1.06
	t =	6.74***	
Focus	$\overline{X}$	2.34	2.09
	S.D.	.54	.63
	t =	2.93**	
Organization	$\overline{X}$	1.95	1.70
	S.D.	.62	.78
	t =	2.41**	
Support	$\overline{X}$	1.94	2.02
	S.D.	.67	.71
	t =	.87	
Mechanics	$\overline{X}$	2.35	2.29
	S.D.	.65	.74
	t =	.65	
Total Score	$\overline{X}$	10.50	9.15
	S.D.	2.66	2.93
	t =	3.37**	
	N =	111	N = 89

Table 11

Correlations Between Expository and Narrative Multiple Choice Scores

	Expository Subscales				
Narrative Subscales	Focus	Organization	Support	Expository Total	Test Total
Focus	.47	.32	.41	.50	.64
Organization	.30	.48	.33	.49	.73
Support	.43	.35	.49	.53	.64
Narrative Total	.50	.52	.51	.65	.88
Test Total	.64	.73	.63	.86	
N	241				

Response Mode Comparisons

The second major research question of the study asked, "What is the relationship among writing scores across response modes?" This question was addressed at two levels of analysis. The first analysis presents simple correlations between the subscales and response modes. The second performs a multi-trait multi-method analysis.

Correlations. Table 12 presents for each of the four test conditions the correlations between the three response modes: 1) between the two essay ratings, 2) between the paragraph and essay ratings, 3) between the multiple choice scores and essay ratings and 4) between the multiple choice scores and paragraph ratings. The correlations are reported both for total scores ahd for the separate subscales.

(Insert Table 12 here)

Relationship of paragraphs to essays. Examination of the relationship between paragraphs and essays on the total scores in the Same Genre condition reveals that the correlation between paragraph and essay ratings are comparable to correlations between the two essay ratings. The exception is the low correlation between the Narrative paragraph (NP) and Narrative Violence (NV) essay in the ND, NV condition. The NV essay scores fluctuated quite dramatically in all test conditions; students seemed to have difficulty with the task. As a consequence, any condition involving NV essays is affected.

In the Different Genre conditions, the total score correlations between paragraphs and essays when the paragraph is of the same genre are comparable to or higher than correlations between the two (different genre) essay scores. This is not surprising, since the two essays are of different genres. Interestingly, in the ND, EV conditions correlations between essay and paragraph total scores are higher than the between-essay correlations, even when the paragraph is

TABLE 12

Within-trait Multimethod Comparison of Correlations Between Response Modes

SAME GENRE

Condition 1: ED, EV
Response Modes (Methods)

Subscales (traits)	ED, EV	PE, ED	PE, EV	MCE, ED	MCE, EV	MCE, PE
General Impression	.56	.65	.47	x	x	x
Focus	.43	.40	.20	.26	.29	.40
Organization	.42	.51	.37	.28	.03	.42
Support	.27	.37	.36	.33	-.09	.30
Mechanics	.58	.77	.64	x	x	x
Total	.60	.75	.61	.43	.23	.58
N	40	40	40	40	40	55

Condition 2: ND, NV
Response Modes (Methods)

Subscales (traits)	ND, NV	PN, ND	PN, NV	MCN, ND	MCN, NV	MCN, PN
General Impression	.39	.08	-.21	x	x	x
Focus	.68	.50	.46	.38	.46	.49
Organization	.42	.13	-.07	.15	.34	.16
Support	.28	.43	.21	.25	.18	.42
Mechanics	.68	.54	.60	x	x	x
Total	.55	.32	.13	.44	.49	.49
N	39	39	39	39	39	64

ED = Expository essay, Drugs
EV = Expository essay, Violence
ND = Narrative essay, Drugs
NV = Narrative essay, Violence

PE = Expository Paragraph
PN = Narrative Paragraph
MCE = Multiple Choice, Expository scores
MCN = Multiple Choice, Narrative scores

(continued)

Table 12

DIFFERENT GENRE (continued)

Condition 3: ND, EV
Response Modes (Methods)

Subscales (traits)	ND, EV	PE, ND	PE, EV	PN, ND	PN, EV	MCE, ND	MCE, EV	MCN, ND	MCN, EV	MCE, PE	MCE, PN	MCN, PE	MCN, PN
General Impression	.33	.48	.59	.40	.50	x	x	x	x	x	x	x	x
Focus	.41	.69	.36	.32	.30	.32	.32	.50	.06	.03	-.08	.33	.28
Organization	.20	.55	.46	.46	.29	.33	.18	.23	.16	.11	.27	.12	.27
Support	.07	.35	.39	.48	.49	.23	.12	.23	.17	.22	.34	.21	.54
Mechanics	.63	.61	.47	.61	.83	x	x	x	x	x	x	x	x
Total	.41	.71	.57	.60	.53	.44	.32	.42	.35	.41	.37	.47	.41
N	40	40	40	40	40	40	40	40	40	31	23	31	23

Condition 4: NV, ED
Response Modes (Methods)

Subscales (traits)	NV, ED	PE, NV	PE, ED	PN, NV	PN, ED	MCE, NV	MCE, ED	MCN, NV	MCN, ED	MCE, PE	MCE, PN	MCN, PE	MCN, PN
General Impression	-.08	-.03	.31	.23	.10	x	x	x	x	x	x	x	x
Focus	.37	.03	.09	-.06	.12	.06	.20	.13	.35	.29	.48	.26	.30
Organization	-.10	-.12	.11	-.21	.36	-.10	.35	-.06	.26	.31	.22	.44	.31
Support	.38	.27	.39	.32	.17	-.01	.26	-.04	.02	.23	.26	.13	.29
Mechanics	.50	.51	.42	.46	.37	x	x	x	x	x	x	x	x
Total	.22	.13	.28	.53	.60	.00	.52	.10	.36	.35	.39	.49	.39
N	54	54	54	54	54	54	54	54	54	34	32	34	32

ED = Expository essay, Drugs
EV = Expository essay, Violence
ND = Narrative essay, Drugs
NV = Narrative essay, Violence

PE = Expository Paragraph
PN = Narrative Paragraph
MCE = Multiple Choice, Expository scores
MCN = Multiple Choice, Narrative scores

in a different discourse mode.

The relationship between paragraphs and essays are more varied when ratings on the separate subscales are examined. Discussion of response mode comparisons for the individual subscales will be treated in more depth in the multi-trait (subscale), multi-method (response mode) analysis presented below.

Relationship of multiple choice tests to essays. In the Same Genre conditions, the correlations between the multiple choice same-genre subscore and essay ratings are lower in the expository comparisons but comparable or higher in the narrative conditions. The greater fluctuations of all students' scores on the narrative constructed response tasks (essays or paragraphs) influence these results. In the Different Genre conditions the correlations between the multiple choice scores and essay scores are generally lower than the constructed response correlations. It should be remembered that the multiple choice test was criterion-referenced, and the resulting restricted variability characteristic of mastery tests may well be partially responsible for the low correlations between the multiple choice test and any of the other measures. Also, the constructed response scores are based on human judgment ratings and are different in kind from the objectives scores.

Relationship between multiple choice test and paragraph. The correlations between the multiple choice test and the paragraph are roughly comparable to the analogous correlations between the multiple choice test and the essays. An exception again involves the comparisons relating the NV essays.

Comparative difficulties of the different response modes. Another way of viewing the comparability of the response mode issue is to compare students' level of performance on the different modes. Table 13 transforms the mean score information reported separately for each response mode (Tables 5, 6, & 9) to percentages of total score. It can be seen that average percentages are roughly comparable on essay and paragraph scores, ranging from 46% to 58%, while performance on multiple choice items is much higher (88% and 82%). These findings are in accord with the notion that recognition tasks are easier than production tasks and may even be considered enroute skills. However, the difference between the bases of the scores may be a factor as well.

(Insert Table 13 here)

Relationships among writing skills and response modes. To examine the relationships between writing skills presumed to be measured by CSE subscales and their hypothesized differential application in different response modes, a multitrait-multimethod analysis was conducted to address the following questions:

1. Do the CSE subscales exhibit:
   a. empirical independence, from which we can infer their conceptual distinctiveness and their potential utility for diagnostic purposes?
   b. empirical homogeniety across measurement methods, such that each subscale relates to its underlying trait in an invariant manner?
2. Taking as a criterion for subscale content measures derived from full length writing samples (i.e., essays), how do paragraph and multiple choice measures compare with respect to:
   a. the discriminant validity of the information they provide?
   b. their degree of relationship to the underlying trait they purport to measure?

Table 13

Percent Correct on Essays, Paragraph and Multiple Choice Tests

Subscales (traits)	ED	EV	ND	NV	ND	EV	NV	ED	PE	PN	MCE	MCN
	%	%	%	%	%	%	%	%	%	%	%	%
General Impression	55	50	32	22	36	47	52	24	48	27	-	-
Focus	61	57	55	55	57	55	58	57	59	52	92	90
Organization	54	50	47	41	47	49	52	38	49	43	81	66
Support	59	60	61	55	59	57	57	53	49	51	91	88
Mechanics	59	59	64	61	57	51	61	59	59	57	-	-
Total	58	54	52	47	51	52	56	45	53	46	88	82

ED = Expository essay, Drugs
EV = Expository essay, Violence
ND = Narrative essay, Drugs
NV = Narrative essay, Violence

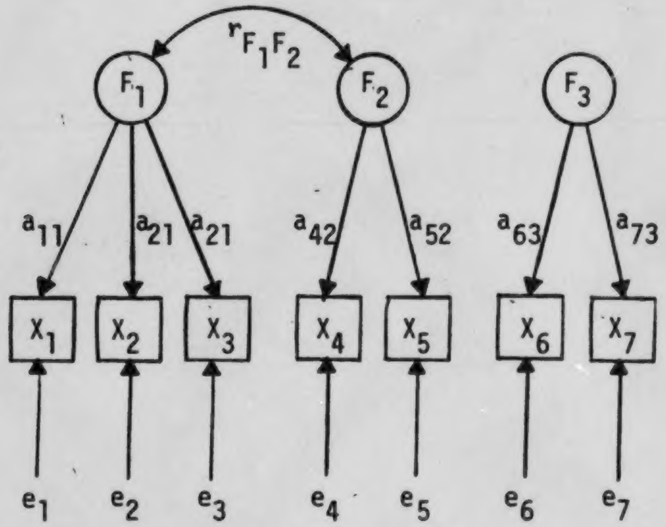
PE = Expository Paragraph
PN = Narrative Paragraph
MCE = Multiple Choice, Expository scores
MCN = Multiple Choice, Narrative scores

3. Are there particular subscale-response mode combinations that are of especially good or poor quality from a measurement standpoint?

The method of analysis of covariance structures (Joreskog, 1970) was employed to explore these related questions. Covariance structure modeling permits the test of hypotheses specified in advance concerning the underlying structure and/or sources of variation that may be operating in a given set of data. For example, models can be formulated that assess whether and to what extent the information about writing skill derived from paragraph and multiple choice response modes is equivalent to that obtained from essays. In addition, the analyses can address questions about the distinctiveness and/or diagnostic utility of the five subscales generated by the CSE scoring rubric. The following sections present the results of several such analyses, preceded by a brief introduction to their logic and interpretation.

The LISREL Model

All of the covariance structure models discussed in this section are special cases of the general LISREL model for the analysis of linear structural relations developed by Joreskog (1973, 1977; Joreskog & Sorbom, 1978). LISREL may be parameterized to carry out restricted maximum likelihood factor analysis and to estimate the parameters of structural regessions involving the resulting factors. LISREL models may treat parameters in one of three ways: (a) as *free* parameters to be estimated by the computer program; (b) as *fixed* parameters specified in advance to equal some fixed number; and (c) as *constrained* parameters to be estimated by the program subject to the constraint that they equal other estimated parameters. The following illustrations of a simple factor analysis model presents these features of LISREL, along with a heuristic device for displaying particular LISREL models.


$r_{F_1F_2}$
F_1
F_2
F_3
a_{11}
a_{21}
a_{21}
a_{42}
a_{52}
a_{63}
a_{73}
X_1
X_2
X_3
X_4
X_5
X_6
X_7
e_1
e_2
e_3
e_4
e_5
e_6
e_7

factor analysis equations

$$X_1 = a_{11}F_1 + e_1$$

$$\vdots$$

$$X_7 = a_{73}F_3 + e_7$$

factor loading matrix

$$\begin{array}{cccc} & F_1 & F_2 & F_3 \\ \hline X_1 & a_{11} & 0 & 0 \\ X_2 & a_{21} & 0 & 0 \\ X_3 & a_{21} & 0 & 0 \\ X_4 & 0 & a_{42} & 0 \\ X_5 & 0 & a_{52} & 0 \\ X_6 & 0 & 0 & a_{63} \\ X_7 & 0 & 0 & a_{73} \end{array}$$

factor correlation matrix

$$\begin{array}{cccc} & F_1 & F_2 & F_3 \\ \hline F_1 & 1.0 & & \\ F_2 & r_{F_1F_2} & 1.0 & \\ F_3 & 0.0 & 0.0 & 1.0 \end{array}$$

The upper portion of the figure shows the path diagram representation for a factor analysis in which three factors are hypothesized to give rise to or "cause" observed variation and covariation in a set of seven variables. Each observed variable is composed of one factor, F_1, F_2, or F_3, and a unique component e_i (where i refers to the variable number). The a's associated with the arrows in the path diagram from the factor to each X-variable correspond to factor loadings; the a's appear as coefficients in the factor analysis equations and as entries in the factor loading matrix in the lower portion of the figure. In the example, the loadings of X_2 and X_3 are constrained to equal one another, as indicated by the fact that the same symbol, a_{21}, is used to refer to their values. The curved double headed arrow connecting factors 1 and 2 symbolizes their correlation, $r_{F_1F_2}$; the absence of such arrows reflects the restriction that the corresponding factors have correlations that are fixed to equal zero, as seen in the factor correlation matrix. The analyses reported below will follow this general procedure and format: Models of special interest will be displayed in path diagram form, and all results will be presented as matrices of factor loadings and factor intercorrelations.

Two other features of the LISREL approach should be mentioned. First, the LISREL computer program outputs a chi-square goodness of fit statistic that indexes the extent to which the specified model provides a reasonable explanation for the structure of the observed data; non-significant chi-square values imply that the hypothesized model cannot be rejected. Second, each free and constrained parameter estimated by the program is accompanied by an estimate of its standard error which can be used for significance testing.

Multitrait-Multimethod (MTMM) Analysis

Campbell and Fiske (1957) proposed the examination of correlation matrices among several "traits" each measured by multiple methods as an informal procedure for investigating what they termed convergent and discriminant validity. They reasoned that *different measures* of the *same trait* should interrelate strongly if the trait possessed convergent validity, and that *different traits* measured by the *same method* should be relatively independent of one another if the traits possessed discriminant validity. Since their original paper, a number of authors have proposed formal analysis models for examining MTMM matrices (Golding & Seidman, 1974; Jackson, 1969; Joreskog, 1974; Werts, Joreskog & Linn, 1972; Werts & Linn, 1970; Yates, 1979). Each of these models is factor analytic in nature, differing only in the type and number of assumptions made; the aim of each is to distinguish two general sources of underlying variation in data organized into an MTMM matrix, trait variance and method variance. Trait variance is captured in trait factors fitted to the data, while method variance is treated differently depending on the particular model employed (the LISREL model specifies method factors in addition to the trait factors). In this framework questions of convergent and discriminant validity become hypotheses about the existence and independence of trait factors. In LISREL, more refined hypotheses can be examined concerning the relative contributions of trait and method factors to the explanation of the observed data.

The analyses treated the five subscales generated by the CSE scoring rubric as "traits," and the three response modes (essay, paragraph and multiple choice) as "methods."

Variable Definition and Data Preparation

Four separate measures were taken on each subject in the cross-genre study, two essays, one paragraph, and a multiple choice test. Superimposed on this basic four-condition design were two additional experimental conditions, genre and topic. The full design has topic varying systematically only within the two essay conditions, with which its two levels (drugs and violence) are fully crossed. Genre, on the other hand, is nested within subject groups for the essay and paragraph response mode conditions, but is fully crossed for the multiple choice condition.

In order to profitably examine the questions identified above within the MTMM analysis required multiple measures of each variable of interest, e.g., Mechanics measured by essay, Mechanics measured by paragraph, etc. In the context of the cross-genre study data, this means that the only distinctions among variables designated as relevant (for the purpose of the MTMM analysis) relate to the subscale they measure and the method of measurement employed to obtain them.[1] This is equivalent to saying we wish to "pool" over genre and topic, whether they are crossed or nested with response mode, before conducting the MTMM analyses. Obviously, variation due to genre is not irrelevant, since one of the aims of the study was to examine the effects of this very factor. The analyses reported here, however, provisionally placed genre lower in priority.

---

[1] It should be noted that the current focus on subscale content and response mode is directly in line with the additional work promised in the May 1979 Interim Report (cf. Quellmalz, 1979, p. 34).

Earlier analyses of the cross-genre study data revealed that the distributional properties of measures obtained in different genre-topic combinations (e.g., means and variances) varied, and that, in some cases, corresponding correlations from different treatment conditions were unequal (Quellmalz, 1979). The MTMM analyses control for these differences to allow examination of "pure" trait and response mode variations. Since the objective of the MTMM analysis is to shed light on the underlying dimensions of individual differences that give rise to the observed matrix of intercorrelations, the effects of distributional differences among conditions can be likened to those of tests of unequal length or difficulty, and may be adjusted by standardizing the scores within conditions. The between-treatment condition differences in correlations (e.g., $r_{GI(NV)xGI(ED)} \neq r_{GI(EV)xGI(ED)}$), on the other hand, may in fact signal real differences in patterns of relationships under different genre-topic combinations. Such differences may be due to a number of things, ranging from the apparent difficulty subjects had in producing narrative writing to sampling differences between genre-treatment groups. In any event, the analyses reported herein pool over subject groups, setting variables derived in different genre and topic conditions to reflect (after within-condition standardization) subscale content and response mode effects.

The MTMM required several transformations of the data. Table 14 shows the configuration of data for the four groups of subjects in the study. For essay and paragraph response modes half of the total sample wrote under each genre-topic combination manipulated in the study. For example, group 1 and group 4 each wrote an expository essay on the topic of drugs, this being designated as the first essay for both groups;

TABLE 14

Study Design and Data Transformation

	Essay			
	Drugs	Violence	Paragraph	Multiple Choice
Study Sample				
Group 1	EE_{11}	EE_{21}	PE_1	MCE_1+MCN_1
Group 2	EN_{22}	EN_{12}	PN_2	MCE_2+MCN_2
Group 3a	EN_{23}	EE_{13}	PE_3	MCE_3+MCN_3
Group 3b	EN_{23}	EE_{13}	PN_3	MCE_3+MCN_3
Group 4a	EE_{24}	EN_{14}	PE_4	MCE_4+MCN_4
Group 4b	EE_{24}	EN_{14}	PN_4	MCE_4+MCN_4
Standardization Sample	Drugs	Violence	Paragraph	Multiple Choice
Stage 1				
Exposition	$EE_{11}+EE_{24}$	$EE_{21}+EE_{13}$	$PE_1+PE_3+PE_4$	MCE_i (i=1 to 4)
Narration	$EN_{22}+EN_{23}$	$EN_{12}+EN_{14}$	$PN_2+PN_3+PN_4$	MCN_i (i=1 to 4)
Stage 2	Essay 1	Essay 2	Paragraph	Multiple Choice
	$EE_{11}+EN_{12}$ $+EE_{13}+EN_{14}$	$EE_{21}+EN_{22}$ $+EN_{23}+EE_{24}$	PE_i+PN_j (i=1,3,4) (j=2,3,4)	MCE_i+MCN_i (i=1 to 4)

EE= Expository Essay
EN= Narrative Essay
PE= Expository Paragraph
PN= Narrative Paragraph
MCE= Exp. Mult. Choice
MCN= Narrative Mult. Choice

similarly, groups 2 and 3 wrote narrative essays on drugs, again, their first in both cases. The first two groups of subjects wrote a paragraph in the same genre as their essays, while subjects in the third and fourth groups were randomly assigned paragraphs in one of the two genre. The scores subjects received for their performance in essay and paragraph modes were z-transformed on the basis of the distribution of scores for all subjects writing in the same genre-topic condition. Thus, expository essay on drugs scores (EE_{17} & EE_{24}) were z-scored using the mean and standard deviation from the combined sample created by merging the first essay from group 1 with the first essay from group 4. After all essay and paragraph scores were transformed in this manner, the new scores were restandardized across the total group of subjects to yield the scores used in analysis. Note that a different genre-topic combination is treated as "essay 1" (and essay 2) for each group of subjects (see the "Stage 2 Standardization" portion of the Table).

For the multiple choice response mode, items were administered that were designed to be related to only three of the five CSE subscales, Focus (Main Idea), Organization and Support. Items dealing with both genres were given to all subjects. As for the essay and paragraph scores, within-genre z-transformations were applied to the multiple choice scores, except that the standardization sample was the total sample, as all subjects were administered all items.

The final set of variables entered into the MTMM LISREL analyses consisted of 18 scores for each subject: the five CSE scales x the three constructed response "methods" (essay 1, essay 2, and paragraph), plus the three scores derived from the multiple choice test. A total of 148 subjects had complete data on all 18 scores, and the correlation matrices used in analysis were computed on this group.

Analysis Results

As a partial exploration into the nature of the relationship between the set of writing scores considered in the MTMM analyses and other achievement and verbal ability measures, a descriptive correlational analysis was conducted. Correlations were computed between the writing scores and the test scores listed in Table 15 for all subjects included in the MTMM analyses with scores on any of the tests. Of the 148 subjects eligible for the analysis, from 39 to 92 were found to have test scores for one or more of the tests, and the correlations for these subjects are contained in Table 16 along with the test score means and standard deviations.

TABLE 15

Supplementary Test and Performance Measures

for Cross-Genre Study Subjects

(MMTM Analysis Sample)

Differential Aptitude Test - Verbal	(ADTP)
Preliminary Scholastic Aptitude Test	(PRESAT)
Scholastic Aptitude Test	(SAT)
Fall, 1978 English Grades	(FALLGR)
Spring, 1979, English Grades	(SPRGR)
Short Form Test of Academic Aptitude	(VERBP)

TABLE 16

Descriptive Statistics for Test Scores and Intercorrelations Between Test Scores and Writing Measures

	APTP	PRESAT	SAT	FALLGR	SPRGR	VERBP
GIE_1	.421**	.365**	.205	.141	.083	.220
GIE_2	.380**	.232*	.469**	.184	.194	.320*
GIP	.387**	.169	.207	.175	.269*	.422**
FE_1	.351**	.362**	.369*	.132	.200	.478**
FE_2	.372**	.498**	.499**	.323**	.187	.508**
FP	.334**	.184	.363*	.216*	.210	.309
FMC	.406**	.345**	.313*	.199	.146	.373*
OE_1	.321**	.317**	.333*	.170	.029	.282
OE_2	.369**	.300**	.601**	.166	.121	.466**
OP	.417**	.279*	.113	.211*	.288**	.394*
OMC	.500**	.448**	.469**	.137	.392**	.474**
SE_1	.204	.191	.355*	.095	.128	.396*
SE_2	.321**	.269*	.274	.156	.141	.052
SP	.287**	.199	.347*	.193	.149	.486**
SMC	.540**	.481**	.139	.228*	.357**	.288
ME_1	.356**	.420**	.470**	.076	.244*	.283
ME_2	.473**	.432**	.487**	.207*	.381**	.438**
MP	.435**	.293**	.538**	.179	.196	.365*
$\overline{X}$	63.94	443.58	513.26	3.93	4.09	75.08
SD	38.56	95.03	129.30	.80	.69	20.90
n	92	81	43	92	91	39

All correlations in the table are positive, with 68 of the 108 coefficients reaching statistical significance ($p=.05$ or less). School grades (FALLGR & SPRGR) show the weakest relations to writing scores, with 24 of the 36 coefficients between these two sets failing to differ significantly from zero. Of the 16 remaining non-significant correlations, 11 are between writing scores and test variables for which the n's are considerably reduced ($n_{SAT}=43$, and $n_{VERBP}=39$). In general then, there appears to be a moderate level of covariation between the writing scores and other aptitude and ability measures.

Table 17 presents the median intercorrelations between the set of test scores on one hand and the CSE subscales and response modes on the other.

TABLE 17

Median Intercorrelation Between Test Scores

and CSE Subscores and Response Modes

	Subscale
GI	.22
F	.34
O	.32
S	.25
M	.37

	Response Mode
Essay	.32
Paragraph	.28
Multiple Choice	.36

For example, the figure of .22 for General Impression is the median correlation across the rows in the previous table containing coefficients for any of the General Impression measures, i.e., the first three rows. For response mode, the medians are across all subscales measured by that mode. The medians in each case are between .22 and .37. General Impression and Support appear to be the least related to the set of test scores, while Mechanics is most highly related to them. Of the response modes, the multiple choice format produces scores most strongly related to the test scores, lending support to the view that writing skills are less distinctively measured through this response mode. Because of the differences in sample sizes and in the content of the different tests, it would be speculative to probe further into the meaning of these correlations. Planned secondary analyses will attempt to examine more formally the interrelations among the measures of writing and general ability.

Generally, the CSE writing measures are moderately related to other indices of students' aptitude and achievement, with only mild differences from one subscale to the next in the degree of this relationship. The same is true for the three response modes in which writing data were gathered in the cross-genre study, with response mode differences providing mild support for the hypothesis that writing is less well measured in the multiple choice format.

The MTMM analyses begin by considering the data for the "essay 1" and "essay 2" methods only, examining the ten scores defined for these two conditions, two measures each of General Impression (gie_1 and gie_2), Focus (fe_1 and fe_2), Organization (oe_1 and oe_2), Support (se_1 and se_2), and Mechanics (me_1 and me_2). The model specified for these variables includes five "trait" factors (one for each subscale) and two "method" factors (one for each essay). Figure 1 illustrates Model I, and the LISREL estimates of the free and constrained model parameters (along with their standard errors in parentheses) are contained in Table 18. The figure shows Model I allowing the trait or subscale factors to be freely intercorrelated, while the method factors are specified to be uncorrelated with each other and with the subscale factors. Leaving the

FIGURE 1

Path Diagram for Model I

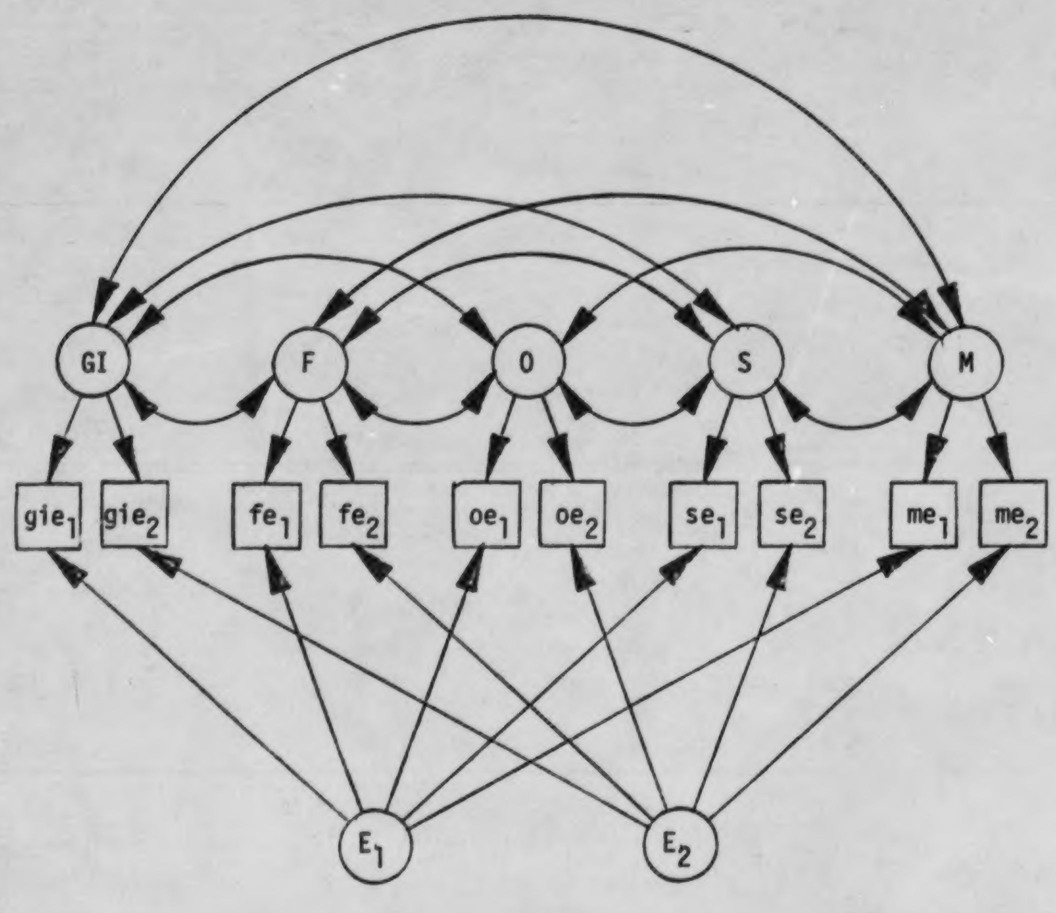

GI
F
O
S
M
gie_1
gie_2
fe_1
fe_2
oe_1
oe_2
se_1
se_2
me_1
me_2
E_1
E_2

TABLE 18

LISREL Estimates for MTMM Model Involving 5 Traits and 2 Methods

LAMBDA	GI	F	O	S	M	E_1	E_2
gie$_1$	.550 (.077)	0	0	0	0	.776 (.079)	0
gie$_2$	.550 (.077)	0	0	0	0	0	.657 (.087)
fe$_1$	0	.641 (.069)	0	0	0	.209 (.090)	0
fe$_2$	0	.641 (.069)	0	0	0	0	.428 (.088)
oe$_1$	0	0	.521 (.081)	0	0	.726 (.085)	0
oe$_2$	0	0	.521 (.081)	0	0	0	.769 (.084)
se$_1$	0	0	0	.557 (.077)	0	.498 (.087)	0
se$_2$	0	0	0	.557 (.077)	0	0	.408 (.094)
me$_1$	0	0	0	0	.770 (.062)	.237 (.081)	0
me$_2$	0	0	0	0	.770 (.062)	0	.182 (.080)

PSI	GI	F	O	S	M
GI	1.0				
F	.721 (.115)	1.0			
O	.916 (.061)	.849 (.113)	1.0		
S	.907 (.100)	.791 (.113)	.792 (.119)	1.0	
M	.772 (.101)	.723 (.087)	.684 (.116)	.661 (.110)	1.0

CHI Square with 20 df = 26.915, *p* = .138.

trait intercorrelations free to be estimated reflects our expectation that the five components of writing ability tapped by the subscales are not independent of one another. The restrictions on the method factor correlations, on the other hand, reflect the hypothesis that they act as independent additive components in the explanation of the observed scores. In addition, the matrix of factor loadings (hereafter referred to as lambda) in the table reveals that we have constrained the loadings of each pair of subscale measures on their corresponding trait factor to equal one another. These constraints are equivalent to a test of the hypothesis that subscale scores from different essays will exhibit the same degree of relationship to the trait factor they measure. The model as a whole cannot be rejected; the chi-square goodness of fit test yields a probability of .138 (ns), suggesting that the model provides an adequate account for the observed data.

Loadings of the essay variables on their corresponding trait factors are all large in magnitude and highly significant, ranging from a low of .521 for Organization to a high of .77 for Mechanics. Except for General Impression and Organization, the loadings of subscale scores on method factors are moderate. One interpretation for the relatively high concentration of method variance in both gie_1 and gie_2 and oe_1 and oe_2 is that residual trait variance that is method specific is shared by these two subscales. This could be the case if raters depend on their impression of the organization of a given writing product, more so than on other characteristics of it, in formulating their general impression rating. Turning to the matrix of factor intercorrelations (hereafter called psi),* we see that the estimates of the relations among the trait factors are all quite high, ranging from a low of .661 for the correlation between Mechanics and Support to a high of .916 between General Impression and Organization. The Mechanics factor appears to be the most independent of the set.

---

* Since method factors are restricted to be uncorrelated with one another and with trait factors, the table omits the corresponding portions of the psi matrix which contains only _fixed_ parameters.

It should be noted that, although the trait intercorrelations are large, their independence from the method and error components in the data (as specified in the model) constitutes a comprehensive form of "disattenuation," making it unlikely that their true values are any higher than those estimated under Model I. Notably, restricting the method factors so that they are independent of one another is not inconsistent with the data; this is to be interpreted as evidence that the primary sources of covariation among scores derived from different essays are capture in the trait factors.

Model II adds paragraph as a method and expands to fifteen the number of variables included in the analysis, by adding the five subscale scores defined for the paragraph response mode. The five trait factors specified in Model I will, under Model II, each have an additional measure of the corresponding trait loading on them (no constraints are placed on these loadings); and there will be a new method factor, Paragraph, to absorb irrelevant covariation specific to this mode of responding. Table 19 presents the results of the LISREL estimation of Model II.

Model II provides an adequate overall fit to the observed intercorrelations (chi-square with 70 df = 79.173, p = .212). This result provisionally supports the hypothesis that the scores generated by application of the CSE scoring rubric to paragraph-length writing samples can

TABLE 19

LISREL Estimates for Model II

LAMBDA	GI	F	O	S	M	E_1	E_2	P
gie_1	.560 (.067)	0	0	0	0	.792 (.069)	0	0
gie_2	.560 (.067)	0	0	0	0	0	.653 (.080)	0
gip	.531 (.089)	0	0	0	0	0	0	.728 (.073)
fe_1	0	.618 (.066)	0	0	0	.212 (.080)	0	0
fe_2	0	.618 (.066)	0	0	0	0	.415 (.080)	0
fp	0	.488 (.087)	0	0	0	0	0	.456 (.079)
oe_1	0	0	.524 (.070)	0	0	.699 (.076)	0	0
oe_2	0	0	.524 (.070)	0	0	0	.756 (.077)	0
op	0	0	.436 (.093)	0	0	0	0	.812 (.074)
se_1	0	0	0	.543 (.067)	0	.511 (.077)	0	0
se_2	0	0	0	.543 (.067)	0	0	.403 (.085)	0
sp	0	0	0	.648 (.089)	0	0	0	.415 (.078)
me_1	0	0	0	0	.778 (.061)	.205 (.068)	0	0
me_2	0	0	0	0	.778 (.061)	0	.138 (.071)	0
mp	0	0	0	0	.728 (.077)	0	0	.214 (.070)

PSI	GI	F	O	S	M
GI	1.0				
F	.746 (.085)	1.0			
O	.938 (.039)	.876 (.081)	1.0		
S	.866 (.064)	.889 (.073)	.871 (.073)	1.0	
M	.802 (.070)	.787 (.071)	.767 (.085)	.697 (.078)	1.0

CHI-Square with 70 df = 79.173. *p* = .212.

be interpreted as measuring the same underlying content as the scores derived from full length essays. Inspection of the lambda matrix shows that the loadings for paragraph subscale scores on their associated trait factors are of substantial magnitude in each case, and that the loadings on the paragraph factor follow the same general pattern as for the two essay method factors. With one exception, the paragraph variables appear to relate to trait factors less strongly than do the essay scores. The exception is an interesting one: "sp" provides a clearer definition of the Support factor (i.e., the loading on "S" is higher for sp than for se_1 and se_2). This would seem to suggest that the task of judging the use of support is carried out more accurately in the context of a single paragraph than it is in longer writing samples;* a test of this hypothesis, however, would require multiple measures of the sp variable.

As in Model I, the trait intercorrelations in Psi are all quite large, indicating considerable interdependence among the subscales. Again, Mechanics exhibits lower levels of relationship to the other subscales.

Comparison of Models I and II reveals two main differences. First, there is some instability in the size of the essay variables' loadings on the associated trait factors as we move from the first to the second model. This leads to the interpretation that the factors composed of both essay and paragraph variables do not measure precisely the same content as factors composed of essay variables only. Second, the estimates of trait intercorrelations in Model II are slightly greater in magnitude than the corresponding Model I estimates. Thus, although the inclusion of paragraph scores may have broadened the content of the factors, it seems also to have diminished their distinctiveness. Depending on one's a priori notions about the comparative validity of essay and paragraph data, Model II may be moving us closer to or further away from the true state of affairs. While the differences between the models are relatively small, we will examine this issue in more detail in the context of Model IV.

---

* Partial support for this interpretation can be found in the comparison of the variance components for "topic" which is confounded with condition in the G-studies run on the essay ratings (Quellmalz, 1979, Appendix G). Of the subscales with specific content (i.e., excluding General Impression), the largest Condition/Topic components are found for the Support scores.

The third MTMM analysis builds on the previous two by adding the three scores derived from the multiple choice items administered to study subjects. Recall that only items analogous to the Focus, Organization and Support subscales were included in the multiple choice test. Model III differs from Model II, then, by the specification of trait loadings for these three subscores and the addition of a multiple choice method factor. Figure 2 displays the path diagram for Model III (and Model IV); and Table 20 the LISREL estimates of the model parameters.

As in the first two analyses, Model III provides a reasonably good fit to the data (chi-square with 112 df = 125.163, p=.186), implying that the same 5-trait structure is not violated by the inclusion of the multiple choice scores. The sizes of the trait loadings for the three response modes, as well as the increases in the trait intercorrelations suggest that the trait factors have drifted closer together as a result of adding the multiple choice variables. Thus, while the multiple choice scores apparently share some content with the writing variables to which they are purportedly analogous, they seem also to possess a higher degree of "latent collinearity" (Yates, 1979) in the trait factor space. Whether

FIGURE 2

Path Diagram for Models III & IV

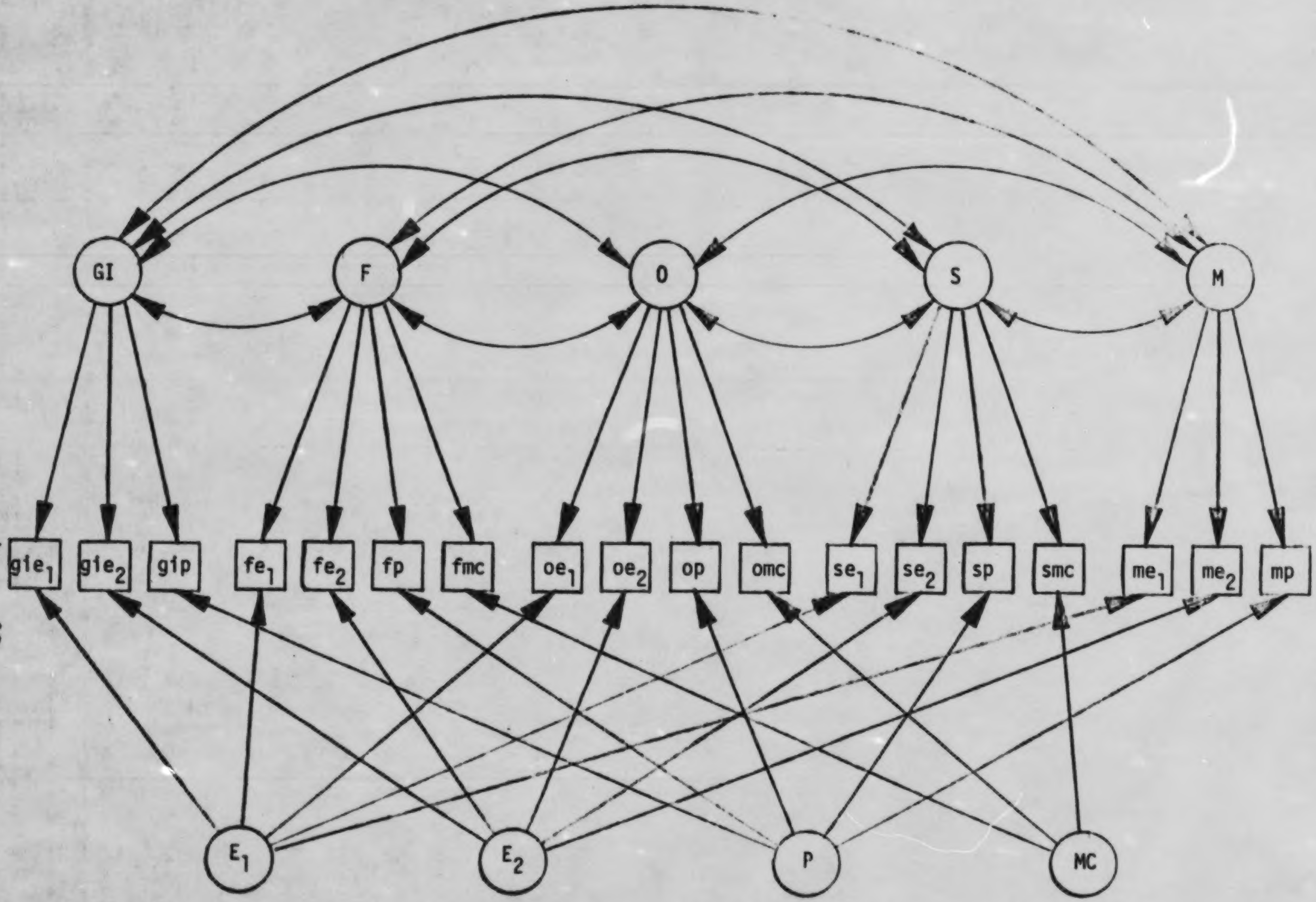

GI
F
O
S
M
gie$_1$
gie$_2$
gip
fe$_1$
fe$_2$
fp
fmc
oe$_1$
oe$_2$
op
omc
se$_1$
se$_2$
sp
smc
me$_1$
me$_2$
mp
E$_1$
E$_2$
P
MC

Table 20: LISREL Estimates for Model III

LAMBDA

	GI	F	O	S	M	E_1	E_2	P	MC
gie_1	.533 (.065)	0	0	0	0	.813 (.067)	0	0	0
gie_2	.533 (.065)	0	0	0	0	0	.677 (.077)	0	0
gip	.586 (.087)	0	0	0	0	0	0	.702 (.073)	0
fc_1	0	.604 (.065)	0	0	0	.239 (.078)	0 0	0 0	0 0
fe_2	0	.604 (.065)	0	0	0	0	.399 (.076)	0	0
fp	0	.535 (.087)	0	0	0	0	0	.423 (.080)	0
fme	0	.512 (.087)	0	0	0	0	0	0	.442 (.124)
oe_1	0	0	.495 (.067)	0	0	.734 (.074)	0	0	0
oe_2	0	0	.495 (.067)	0	0	0	.748 (.075)	0	0
op	0	0	.490 (.090)	0	0	0	0	.780 (.075)	0
omc	0	0	.520 (.090)	0	0	0	0	0	.463 (.126)
se_1	0	0	0	.487 (.066)	0	.531 (.076)	0	0	0
se_2	0	0	0	.487 (.066)	0	0	.420 (.083)	0	0
sp	0	0	0	.636 (.087)	0	0	0	.392 (.079)	0
smc	0	0	0	.458 (.091)	0	0	0	0	.411 (.121)
me_1	0	0	0	0	.772 (.061)	.220 (.067)	0	0	0
me_2	0	0	0	0	.772 (.061)	0	.138 (.069)	0	0
mp	0	0	0	0	.746 (.077)	0	0	.186 (.071)	0

PSI

	GI	F	O	S	M
GI					
F	.789 (.072)	1.0			
O	.933 (.037)	.915 (.062)	1.0		
S	.919 (.057)	.943 (.062)	.953 (.058)	1.0	
M	.816 (.064)	.785 (.065)	.783 (.073)	.766 (.072)	1.0

CHI-SQUARE W/112 df = 125.163, *p* = .186

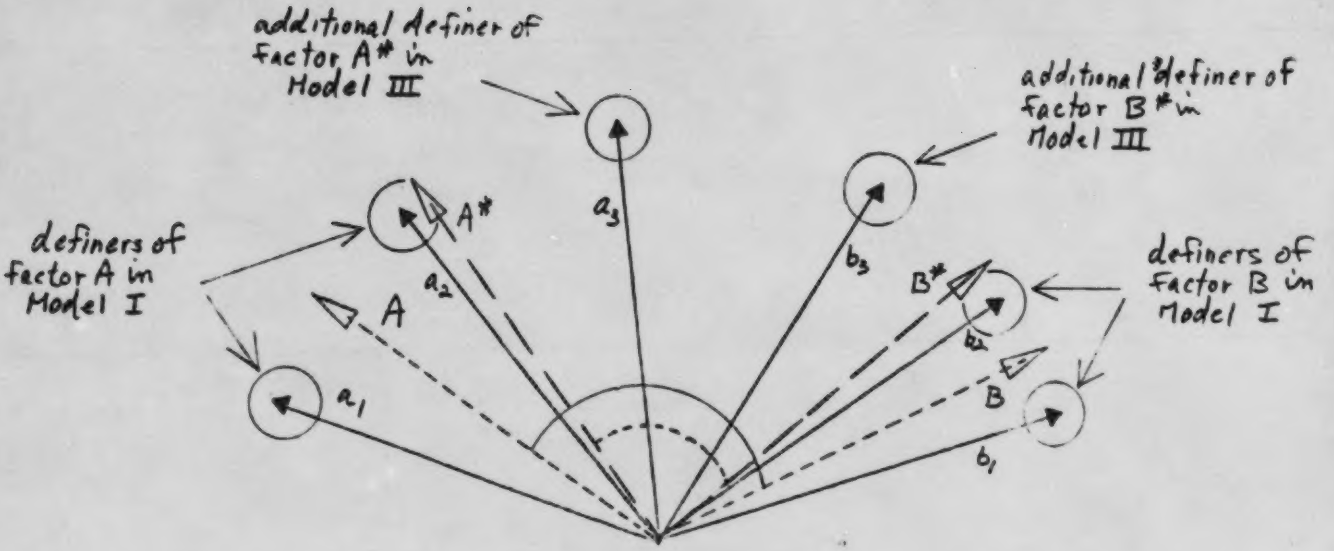
this situation arises because the multiple choice variables are related to writing ability in some non-specific fashion or because all of the variables, but especially the multiple choice scores, share a common dependence on general ability, would require additional analyses which include individual ability test scores. Secondary analyses may pursue this issue using the test data reported in Table 15. In any event, we are more confident here than for Model II in interpreting the increased interdependence among trait factors as an indication that the multiple choice scores possess generally lower validity as distinctive components of writing ability than do measures derived from actual writing samples.

Before presenting Model IV it will be helpful to describe how factor analysis can be interpreted geometrically. Statistically, factor loadings are akin to regression coefficients describing a linear relationship between an underlying factor and an observed variable; geometrically, the size of a factor loading reflects how close in space two lines emanating from a common point, one corresponding to the variable and the other to the factor, are. The closer the line for the factor is to that for the variable, the higher will be the loading. Correlations between factors have the same spatial interpretation. The type of factor analysis employed here restricts the locations of factors through the a priori specification of which variables do and do not load on them. The interpretation of Model III envisioned the spaces among the trait factors becoming smaller, or the lines representing the factors moving toward a common point in the center of the space they defined, as a result of including a set of variables (the multiple choice scores) which, to a great extent than was true for the variables defining the factors in previous models, were themselves located near the center of this space. This reasoning led to the assertion that the multiple choice scores conveyed less distinctive information about writing skill. Figure 3 illustrates the point for an idealized version of two factors from the first and third models. The variables a_1 and a_2 define factor A and the variables b_1 and b_2 define factor B in Model I, and the solid arc corresponds to the correlation between the resulting factors. By adding the variables a_3 and b_3 the Model III estimates of the positions of the new factors A* and B* place them closer together, as indicated by the shorter dashed arc. Specifying that the variables a_3 and b_3, which, while related to the other a's and b's respectively, are also quite highly related to each other, load only on the corresponding A and B factors, the resulting A* and B* factors are more highly correlated than were their analogs from the first model. It can also be noted that a_3 and b_3 have higher loadings on the Model III factors than they would on the Model I factors.

Model IV examines the relationship of the paragraph and multiple choice variables to the set of trait factors defined solely on the basis of the essay variables. The Model IV rests on the assumption that, at least in the case of the multiple choice scores, the essay-only factors presented a clearer picture of the underlying content of the CSE writing scores; Model IV treats those factors as "unmeasured" criterion variables against which to compare scores from the other two response modes. This can be accomplished in LISREL by modifying the specification for Model III in two places. First, instead of estimating trait loadings for the essay variables new specifications fix their values to equal those estimated in Model I. Second, fixing their values at those obtained in the essay-only solution for Model I places a similar constraint on the trait intercorrelations in psi. These two sets of restrictions will ensure that the

FIGURE 3

Latent Collinearity and Factor Collapse


additional definer of
factor A* in
Model III
additional definer of
factor B* in
Model III
definers of
factor A in
Model I
definers of
factor B in
Model I
A*
A
a_1
a_2
a_3
b_3
B*
b_2
B
b_1

trait factors found in Model I will reappear in Model IV. The LISREL estimates of the free parameters in Model IV are contained in Table 21.

The only parameter estimates of direct interest in Table 21 are the trait factor loadings for the paragraph and multiple choice variables. The data indicate near uniform reduction in their magnitude in comparison to the estimates obtained from Model III. This shift does not reduce overall model fit (chi-square with 127 df=136.919, p=.258). In all but one instance paragraph and multiple choice trait factor loadings are lower than the corresponding loadings for essay variables. The one exception is a recurrance of the finding from Model II that the measure of Support derived from a paragraph-length writing sample outperforms the Support measures based on full length essays. On the other hand, Support as measured by multiple choice items seems to reflect relatively little of what is measured in actual writing samples. The remaining two multiple choice scores, mcf and mco, seem to convey a roughly comparable amount of information about subscale content to that contained in a single paragraph.

Table 21: LISREL Estimate for Model IV

LAMBDA

	GI	F	O	S	M	E_1	E_2	P	MC
gie_1	.550	0	0	0	0	.788 (.068)	0	0	0
gie_2	.550	0	0	0	0	0	.656 (.078)	0	0
gip	.520 (.079)	0	0	0	0	0	0	.704 (.020)	0
fe_1	0	.641	0	0	0	.260 (.079)	0	0	0
fe_2	0	.641	0	0	0	0	.381 (.076)	0	0
fp	0	.485 (.083)	0	0	0	0	0	.448 (.078)	0
fmc	0	.477 (.085)	0	0	0	0	0	0	.451 (.117)
oe_1	0	0	.521	0	0	.746 (.074)	0	0	0
oe_2	0	0	.521	0	0	0	.738 (.075)	0	0
op	0	0	.442 (.084)	0	0	0	0	.795 (.073)	0
omc	0	0	.486 (.092)	0	0	0	0	0	.457 (.118)
se_1	0	0	0	.557	0	.515 (.075)	0	0	0
se_2	0	0	0	.557	0	0	.407 (.084)	0	0
sp	0	0	0	.623 (.083)	0	0	0	.406 (.076)	0
smc	0	0	0	.387 (.091)	0	0	0	0	.465 (.122)
me_1	0	0	0	0	.770	.231 (.068)	0	0	0
me_2	0	0	0	0	.770	0	.136 (.069)	0	0
mp	0	0	0	0	.720 (.071)	0	0	.197 (.070)	0

CHI-SQUARE W/127 df = 136.919, p = .258

Table 4: LISREL Estimate for Model IV

LAMBDA

	GI	F	O	S	M	E_1	E_2	P	MC
gie_1	.550	0	0	0	0	.788 (.068)	0	0	0
gie_2	.550	0	0	0	0	0	.656 (.078)	0	0
gip	.520 (.079)	0	0	0	0	0	0	.704 (.020)	0
fe_1	0	.641	0	0	0	.260 (.079)	0	0	0
fe_2	0	.641	0	0	0	0	.381 (.076)	0	0
fp	0	.485 (.083)	0	0	0	0	0	.448 (.078)	0
fmc	0	.477 (.085)	0	0	0	0	0	0	.451 (.117)
oe_1	0	0	.521	0	0	.746 (.074)	0	0	0
oe_2	0	0	.521	0	0	0	.738 (.075)	0	0
op	0	0	.442 (.084)	0	0	0	0	.795 (.073)	0
omc	0	0	.486 (.092)	0	0	0	0	0	.457 (.118)
se_1	0	0	0	.557	0	.515 (.075)	0	0	0
se_2	0	0	0	.557	0	0	.407 (.084)	0	0
sp	0	0	0	.623 (.083)	0	0	0	.406 (.076)	0
smc	0	0	0	.387 (.091)	0	0	0	0	.465 (.122)
me_1	0	0	0	0	.770	.231 (.068)	0	0	0
me_2	0	0	0	0	.770	0	.136 (.069)	0	0
mp	0	0	0	0	.720 (.071)	0	0	.197 (.070)	0

CHI-SQUARE W/127 df = 136.919, *p* = .258

DISCUSSION

The purpose of the study was to examine the relationship among writing performance measures which varied according to discourse and response mode. Results indicated that the relationship between a student's scores on essays written in the same mode of discourse is stronger than the correlation between two essays in different discourse modes. These findings suggest that generalizations about student writing competency should take care to reference the discourse domain rather than the more general domain of writing. Writing skills for Focus, Support, and Mechanics were more stable across genres, while the global General Impression judgment and Organization score seemed to capture genre-specific elements. These data cast doubt on the assumption that a "good writer is a good writer" regardless of task variations. Logically and pedagogically such a position seems untenable. Composition instructors do not teach organization as an undifferentiated construct. Temporal, logical and chronological organization is often taught separately. Rhetoricians and semanticists assert that the structural properties of modes of discourse differ. It follows, then, that students must master these different structures. The present study's data suggest that students' command of different discourse structures vary.

It might have been expected that narrative writing would be easier for students, but it was not the case in this study. Although the narrative topics and directions were derived from domain specifications and pilot tested, students in the study often did not deal with the topic from a narrative perspective at all, or had great difficulty doing so. It is not clear in this study if narrative writing was a less well formed skill for the students or whether they were not comfortable with the two particular narrative topics.

Statistics used to calculate the generalizability coefficients for the essay ratings revealed that the variance components for subject-by-topic interactions ranged from .01 to .87, while in exposition these figures ranged from .00 to .30. Thus, topic variability within discourse mode remains a methodological issue in writing assessment. Further research should examine decision rules that could be employed at the topic pilot testing stage to minimize subject-by-topic interaction. In addition, more refined analyses, such as the LISREL models employed to examine trait and response mode effects, can further explore the relationship of discourse mode to writing performance.

The multitrait multi-method analyses provided a more detailed investigation of several questions concerning the validity of the CSE scoring rubric as a measure of writing performance and concerning the comparative information yield of alternative response modes for assessing writing.

The MTMM analyses suggest first that repeated applications of the CSE scoring rubric to writing samples in fact produce measures that tap the same underlying content. Thus, given multiple measures of each subscale, it is possible to fit a factor analysis model that confirms their hypothesized content. Second, it was found that factors reflecting the content of the five CSE subscales are highly intercorrelated, and this interdependence appears to be present no matter what response mode subjects are assessed in. Further research will continue to investigate the construct validity of scale criteria. The subscale factor intercorrelations, however, are not so high as to belie the possibility that they index distinctive components of writing skill. The Mechanics subscale displayed the greatest degree of independence.

The MTMM analyses also produce information on the extent to which the various subscale-response mode combinations contain "method variance" not related to their substantive content. The specific models tested suggest that scores on the General Impression and Organization subscales contain large method components when measures are taken from constructed responses. A plausible explanation for this finding is that the method factor loadings for these variables are inflated by within-occasion (e.g., a given essay) residual linkages between GI and O brought about by raters' tendency to depend more on Organization than on other specific features in formulating their General Impression rating. The remaining three subscales all were found to contain proportionately larger amounts of content-related variance than method related variance, with Mechanics appearing to be the purest of the three. The patterning of method variance saturation in the five subscales was the same for the three writing samples available for each subject.

An interesting picture of the effects of varying response mode emerged from the analyses. While models can be fitted to the data from all three response modes that confirm the subscale content, the degree of independence of the resulting subscale factors appears to be affected by which response modes are included in the analysis. The most differentiated subscale factor structure is obtained by including only essay variables in the analysis; interdependence among the subscale factors increases with the addition of both paragraph and multiple choice measures. Thus, the effect of shortening the assessment task for the examinee through examination of just paragraph or multiple choice tasks does not simply increase the measurement error. The savings in testing time are obtained also at the cost of clarity and distinctiveness in the information about each of the subscales. When the subscale content factors are located in the variable space so as to maximize their relationship to scores derived from the essay response mode, all other subscale-response mode combinations except one provide weaker substantive information. The one exception is the measure of Support based on paragraph-length writing samples which seems to be superior to the corresponding essay variables in its ability to capture subscale content. It may be that the use of support is less equivocally evaluated in the context of a single paragraph than in an essay containing multiple paragraphs, each of which may suggest a different view of the examinee's ability to provide supporting detail.

The results of the MTMM analyses point clearly to a need to examine data generated by the CSE scoring rubric in connection with other performance data. For example, we were unable to satisfactorily account for the finding that the distinctions among the subscales tend to collapse for measures taken in the paragraph and multiple choice response modes. One interpretation would suggest that scores in these modes are simply more general measures of writing ability, and that for fine-grained distinctions (rather than a quick assessment) one should require subjects to generate full length written products. Alternatively, paragraph and selected response tasks may not bear any direct relationship to writing ability at all (we should exempt Support/Paragraph from this interpretation), and instead, are more akin to verbal ability measures. In this second view, all of the measures considered in our analyses contain some component due to general ability, but the shorter tasks lack the same degree of saturation with additional content specific to the writing domain possessed by the essay scores. Secondary analyses may pursue this important issue with the individual test data reported in Table 15.

The questions examined in this series of MTMM analyses do not exhaust possibilities for applying the technique to other questions related to cross-genre study data, in particular, and writing assessment research in general. The covariance structure modeling framework can explore, for example, the influence of genre on writing performance. Such analyses will necessarily be more tentative, since the nested character of the study design precludes the use of the entire database when looking at genre differences. Certainly covariance structure modelling can permit analyses of writing assessment variables far more powerful than techniques commonly employed.

Another potentially important avenue for additional work would be to address more directly the legitimacy of claims that certain aspects of writing ability are well measured by multiple choice tests. In order to pursue this issue in the context of the cross-genre study, we would need to disaggregate subjects' multiple choice test scores to obtain two measures for each of the three subscales assessed in this mode. It is possible to use the two measures of each trait available in the multiple choice data to apply the approach used in Model IV using the multiple choice scores as the basis for a criterion factor solution. Comparison of such a model with a 3-trait version of Model IV should provide an informative test of the utility of this multiple choice data for assessing writing. We are now exploring the feasibility of carrying out such a test.

In summary, the study highlights the importance of precision in designing, analyzing and reporting writing assessment data. It would seem that the technology of domain-referenced specifications provides a reasonable framework for focusing attention, discussion, assessment and instruction on clearly bounded classes of writing performance.

References

Anderson, R. C. The notion of schemata and the educational enterprise. In Anderson, R. C., Spiro, R. J., & Montague, W. E. (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Earlbaum Associates, 1977.

Braddock, R., Lloyd-Jones, R., & Schoer, L. *Research in written composition*. Champaign, Illinois: National Council of Teachers of English, 1963.

Breland, H. M. A study of college English placement and the test of standard written English. Princeton, N. J.: Educational Testing Service, January, 1977.

Bourne, L. J. *Human conceptual behavior*. Boston: Allyn & Bacon, 1966.

Campbell, D. T. & Fiske, D. W. Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959, *56*, 81-105.

Coffman, W. E. Essay exams. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.), Washington, D. C.: American Council on Education, 1971.

Cooper. C., & Odell, L.(Eds.). *Evaluating writing: Describing, measuring, judging*. Buffalo, N. Y.: State University of New York at Buffalo, 1977.

Diedrich, P. B. *Measuring growth in English*. Champaign, Illinois: National Council of Teachers of English, 1974.

French, J. W. Schools of thought in judging excellence of English themes. 1961 Proceedings of Invitational Conference on Testing Problems. Educational Testing Service, Princeton, N. J., 1962.

Godshalk, F. I., Swineford, F., & Coffman, W. E. *The measurement of writing ability.* New York: College Entrance Examination Board, 1966.

Golding, S. L. & Seidman, E. Analysis of multitrait-multimethod matrices: A two step principal components procedure. *Multivariate Behavioral Research*. 1974, *9*. 479-496.

Jackson, D. N. Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin*. 1969, *72*. 30-49

Joreskog, K. G. A general method for the analysis of covariance structures. *Biometrika*. 1970, *57*, 239-251.

Joreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 1973, 85-112.

Joreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, & P. D. Suppes (Eds.), Contemporary Developments in Mathematical Psychology, Vol. II. San Francisco: W. H. Freeman & Co., 1974, 1-56.

Joreskog, K. G. Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), Applications of Statistics. North-Holland Publishing Co., 1977, 265-287.

Joreskog, K. G. & Sorbom, D. LISREL IV. A general computer program for estimation of linear structural equation systems by maximum likelihood methods. University of Uppsala, Department of Statistics, 1978.

Kincaid, G. L. Some factors affecting quality of students' writing. Unpublished Ed.D. dissertion, Michigan State University, 1953. Cited in R. Braddock, et al, Research in written composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Kinneavy, J. L. A theory of discourse. The Aims of Discourse. Englewood Cliffs, N. J.: Prentice Hall, Inc., 1971.

Lloyd-Jones, R. Primary trait scoring. In C. Cooper, L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Buffalo, N.Y.: State University of New York at Buffalo, 1977.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Pitts, M. The relationship of classroom instructional characteristics and writing in the descriptive narrative mode. Report to the National Institute of Education, November, 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

Quellmalz, E. Interim Report. Defining writing domains: Effects of discourse and response mode. Center for the Study of Evaluation, 1979.

Smith, L. S. Investigation of writing assessment strategies. Report to the National Institute of Education, November 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

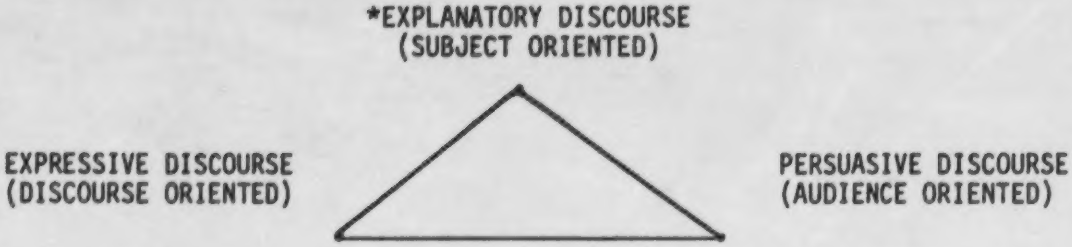
Veal, L. R. & Tillman, M. Mode of discourse variation in the evaluation of children's writing. Research in the Teaching of English, 1971, 5, 37-45.

Werts, C. E., Joreskog, K. G., & Linn, R. L. A multitrait-multimethod model for studying growth. Educational and Psychological Measurement. 1972, 32, 655-678.

Werts, C. E. & Linn, R. L. Path analysis: Psychological examples. *Psychological Bulletin*, 1970, *74*, 193-212.

Winters, L. The effects of differing response criteria on the assessment of writing competence. Report to the National Institute of Education, November 1978. (Grant No. OB-NIE-G-78-0213 to the UCLA Center for the Study of Evaluation.)

Yates, A. Distinguishing trait from method variance in fitting the invariant common factor model to observed intercorrelations among personality rating scales. Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Los Angeles, CA, 1979.


*EXPLANATORY DISCOURSE
(SUBJECT ORIENTED)
EXPRESSIVE DISCOURSE
(DISCOURSE ORIENTED)
PERSUASIVE DISCOURSE
(AUDIENCE ORIENTED)

*Exposition

Lloyd-Jones (1977)

Logical development (i.e., organization of ideas) and the use of explicit evidence traditionally are associated with persuasive discourse--and also are key ingredients of exposition. Persuasion, however, can be "...located more toward the expressive point of the [model's] triangle" (Lloyd-Jones, 1977) in that often its effectiveness depends upon expressions of the writer's voice and appeals of classical rhetoric. Moreover, a piece of persuasive discourse can be judged as effective to the extent that it affects or influences the reader's point of view, attitudes, perceptions.

Herein lies the chief distinction between the type of writing desired in this exercise, exposition, and its cousin, persuasion. Exposition seeks to explain or clarify a subject; it attempts to answer questions about a subject such as:

> "What is it?"
> "What does it mean?"
> "How is it put together?"
> "How does it work?"
> "What was it intended for?"
> "How did it occur or exist?"
> "What is it worth?"
> "What is its importance?"
> "How well does it fulfil its intended function?"
>
> Brooks and Warren, 1961

For example, the topic, "What is the value of________?" can be used to

APPENDIX A

DOMAIN SPECIFICATIONS FOR DEVELOPING EXPOSITORY WRITING SAMPLE TASKS

Objective

Given a writing topic, directions specifying an expository writing purpose and structural elements to include, the student will compose an essay of at least one page (approximately 200 words) within a forty minute time period.

Stimulus Limits

A. Mode of Discourse (Purpose)

The writing assignment should be constructed in a manner that is likely to elicit expository prose. One way to define what is meant by exposition is to consider what it is not: it is not description or narration (although they may be incorporated as elements within an expository piece of writing), nor is it persuasive/argumentative writing. This latter point becomes especially critical when rating criteria are established. The distinction between expository and persuasive discourse becomes clearer, perhaps, when the respective purposes of the writing tasks are taken into account.

Rhetorical Purpose

Lloyd-Jones (1977) has illuminated the importance of rhetorical purpose in the following simple model of discourse:

prompt both persuasive and expository types of writing, but qualities of the final written product likely will be dissimilar due to differing rhetorical aims. For example:

TOPIC: "WHAT IS THE VALUE OF______?"

POSSIBLE AIMS OF PERSUASIVE WRITING	POSSIBLE AIMS OF EXPOSITORY WRITING
. to convince audience of the worth...	. to inform audience about positive aspects of...
. to influence opinions about...	. to explain the writer's reasons for valuing...
. to promote more positive attitudes toward...	. to demonstrate the importance of...
. to prompt action on...	. to increase understanding about...

Judgments regarding the quality of components of an expository piece, i.e., method of development, use of supporting details, use of transitions, thus can be arrayed against the following criterion: To what extent do the components of the essay help promote reader understanding of the subject?

The writing assignment should ask the student to explain or clarify a subject. The student should know that the intent of the essay is to provide knowledge, information, or explanations to a particular audience. These conditions need not prevent the writer from assuming a point of view about a subject (i.e., agreeing/disagreeing; liking/disliking; accepting/rejecting). In fact, a well organized and documented expository piece may be quite persuasive (i.e., it affects reader perceptions) simply because ideas and supporting details have been logically and clearly presented.

B. Topic Limits

The subject (content area) of the writing task should be familiar to the target writers, urban and suburban high school students. The task should allow students to use material drawn from personal experience and popular knowledge, i.e., television, books. The topic might specify a problem or conflict situation which require clarification. However, the topic should not encourage development exclusively through expressions of personal opinion (i.e., "Why do you enjoy______type of music?). Instead, the task should encourage students to express an idea or point of view which can be logically developed through the use of supporting details, such as examples, anecdotes, facts. In addition, the topic should be limited in scope; it should avoid subjects that require extensive definition -- "What is the value of a good education ["good education" must be defined]?"

C. Context

The writing task should specify the rationale for the writing task and the social context to which it is relevant; This may be accomplished by specifying a role the writer is to assume e.g., (a reporter) and the hypothetical intended audience for the essay (e.g., community, school friends, particular agency).

Response Criteria

The essay will be judged according to an analytical scale which includes judgments on overall essay quality, focus, organization, support and mechanics (see Appendix D for expository scale). Essay directions should specify the major criteria on which the essay will be judged.

Format/Directions

1. Task directions should include a brief (2-3 sentence) explication of the topic, including the subject (e.g., Television commercials) the problem or conflict (e.g, how they affect young children).

2. Directions should then specify the rhetorical context (e.g., you have been asked by ______________ to write for______________ purpose for_________________audience).

3. Finally, directions should specify what criteria, elements students should include (e.g., focus, organization etc.).

Sample Item

Many young adults use drugs. In fact, every year these substances hurt the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to explain the problem as it exists in your district. Write an essay to explain why many of today's teenagers abuse drugs.

. Your essay should give an informative explanation, not just opinions.

. Back up your ideas with specific support such as examples, facts, and other details.

. Be sure your essay has a clear main idea and is well organized.

APPENDIX A

Expository Essay Task 1

Many young adults use drugs. In fact, every year these substances hurt the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to explain the problem as it exists in your district. Write an essay to explain why many of today's teenagers abuse drugs.

- Your essay should give an informative explanation, not just opinions.
- Back up your ideas with specific support such as examples, facts, and other details.
- Be sure your essay has a clear main idea and is well organized.

APPENDIX A

Expository Essay Task 2

Many news reports and books assert that today's students do not respect adults. The reports describe instances of discipline problems, even violence, in schools, neighborhoods and in homes. To understand this problem, experts have studied areas such as teaching methods, family life, and the effects of television.

Assume you have been asked by your school board to explain the problem as it affects students in your district. Write an essay in which you explain why many of today's students feel and act this way.

- Your essay should give an informative explanation, not just opinions.
- Back up your ideas with specific support such as examples, facts, and other details.
- Be sure your essay has a clear main idea and is well organized.

APPENDIX B

DOMAIN SPECIFICATIONS FOR DEVELOPING FACTUAL NARRATIVE WRITING SAMPLE TASKS

Objective

Given a topic, directions to write for a factual narrative purpose and a list of narrative elements to include, the student will write a narrative essay of at least 200 words.

Stimulus Attributes

A. Discourse Mode

The essay task should present a topic and employ directions that will clearly elicit narrative prose. The type of narrative writing of interest is termed "factual" narrative or "generalized" narrative. (Brooks and Warren, 1961). In factual narration, the intent is to recount what happened over a period of time. It is a form of what Kinneavy(1971) calls "reference discourse" where the emphasis is on the reality to which reference is made. Similarly, Lloyd Jones (1977) would characterize factual narrative as "subject oriented." Some views on narration are given below:

- Narration is the type of discourse that tells what happened, how it happened.
- Narration gives an impression of movement in time, the sense of witnessing action.
- Narrative action includes time, logic and meaning.

B. Topics Attributes

Topics should be selected which are within the personal experience, immediate or vicarious, of the majority of senior high school students. Topics should be of high interest and present a problem or conflict situation.

C. Format/Directions

Directions should briefly describe a situation, (the topic) direct writers to produce an narrative essay, and present the narrative elements which should be included in the essay.

D. Sample Item

Many young adults use drugs. In fact, every year these substances hurt the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to tell how drugs affect students in your district. Describe what happened over a period of time to one or more people you know who became involved with drugs.

. Your narrative should help others experience how drugs affect students.

. Use specific details and facts to make your account realistic.

. Be sure your narrative has a clear main idea and is well-organized.

APPENDIX B

Narrative Essay Task 1

Many young adults use drugs. In fact, every year these substances hurt the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to tell how drugs affect students in your district. Describe what happened over a period of time to one or more people you know who became involved with drugs.

- Your narrative should help others experience how drugs affect students.
- Use specific details and facts to make your account realistic.
- Be sure your narrative has a clear main idea and is well-organized.

APPENDIX B

Narrative Essay Task 2

Many news reports and books assert that today's students do not respect adults. The reports describe instances of discipline problems, even violence in schools, neighborhoods and in homes.

Assume you have been asked by your school board to tell how students in your district show their lack of respect for adults. Describe what happened over a period of time to one or more people you know who did not respect adults.

- Your narrative should help others experience how these students felt and acted.
- Use specific details and facts to make your account realistic.
- Be sure your narrative has a clear main idea and is well organized.

APPENDIX C

Expository Paragraph

Many young adults use alcohol. In fact, every year alcohol hurts the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to explain the problem as it exists in your district. Write a paragraph to explain why many of today's teenagers abuse alcohol.

- Your paragraph should give an informative explanation, not just opinions.
- Back up your ideas with specific support such as examples, facts and other details.
- Be sure your paragraph has a clear main idea and is well organized.

APPENDIX C

Narrative Paragraph

Many young adults use alcohol. In fact, every year alcohol hurts the lives of thousands of American teenagers by making them emotionally and physically ill.

Assume you have been asked by your school board to tell how alcohol affects students in your district. Write a paragraph to describe what happened to one or more people you know who became involved with alcohol.

- Your narrative should help others experience how alcohol affects students.
- Use specific details and facts to make your account realistic.
- Be sure your narrative has a clear main idea and is well organized.

APPENDIX D

Expository Scale

ELEMENT 1

Impressionistic Rating Procedures

The purpose of Impressionistic Rating is to form a single impression of a piece of writing as to how well it communicates a whole message to the reader. Impressionistic scoring assumes that each characteristic that makes up an essay -- organization of ideas, content, mechanics and so on -- is related to all other characteristics. Impressionistic scoring further assumes that some qualities of an essay cannot easily be separated from each other. In short, the procedure views a piece of writing as a total work, the whole of which is greater than the sum of its parts.

Discerning readers naturally will attend to, or be influenced by, some essay characteristics more than others. In the Impressionistic scoring, however, readers should arrive at a judgment regarding the overall quality of the essay.

For this element, you are being asked to form an overall impression concerning the effectiveness of the essays as examples of expository writing.

The Topics

You will be reading essays on two different topics. Both topics, though, were designed to elicit writing in the expository mode. Some views on exposition are given below:

- Exposition is the kind of discourse that explains or clarifies a subject.
- Exposition seeks to explain or inform through such methods as giving reasons or examples, comparing and contrasting, defining, ennumerating or through a combination of methods.
- Exposition explains why or how.
- Exposition promotes reader understanding of a subject.

General Impression

You are to read each essay quickly; first in order to form an overall impression of its quality. To assign the essay a score, consider the following question: To what extent is the essay an example of effective exposition?

Assign each paper a mark of 1 - 4 using the scale below:

4 = An excellent example of exposition

3 = A good, adequate example of exposition

2 = A minimally adequate example of exposition

1 = A poor example or barely readable example of exposition

APPENDIX D

Expository Scale

ELEMENT 2

Essay Focus/Main Idea

The introduction (if deductively structured) or conclusion (if inductively structured) of the essay clearly indicates the subject and main idea of the whole essay.

4. The introduction (and/or conclusion) of this paper clearly conveys the main idea of the whole essay. It also limits the topic by alerting the reader to the key points covered in the body of the essay.

   Specifically, in the introduction (and/or conclusion):

   a. The subject of the essay is clearly identified.

   b. The main idea of the whole essay is clearly stated or implied.

   c. The topic is clearly limited. That is, key points (e.g., reasons, ideas) or major line(s) of reasoning treated in the essay are identified or summarized.

3. The introduction (and/or conclusion) of this paper conveys the main idea of the whole essay. It sets limits on the topic, but does not clearly suggest how the main idea is developed.

   Specifically, in the introduction (and/or conclusion):

   a. The subject of the essay is clearly identified.

   b. The main idea of the whole essay is clearly stated or implied.

   c. An attempt is made to limit the topic. That is, the number -- or type -- of key points is specified, but there is not clear reference to the substantive issues treated in the body of the essay.

2. The introduction (and/or conclusion) of this paper gives the reader a fairly clear sense of the main idea of the whole essay. However, neither the introduction nor the conclusion help focus -- or bring direction to -- the body of the paper.

   Specifically, in the introduction (and/or conclusion):

   a. The subject of the essay is identified.

   b. The main idea of the whole essay is stated or implied.

   c. No attempt is made to limit the topic.

1. Neither the introduction nor the conclusion is helpful to the reader in obtaining any sense of the main idea of the essay.

   Specifically, in the introduction (and/or conclusion):

   a. The subject of the essay is not clearly identified or there is no reference to the subject

   AND/OR

   b. The main idea of the whole essay is not clearly stated or implied or no reference is made to the main idea, or the reference is confusing.

APPENDIX D

Expository Scale

ELEMENT 3

Essay Organization

The main idea is developed according to a clearly discernible method of organization.

4. The plan by which the essay is structured is readily apparent and consistently applied throughout the essay. Some logical scheme serves as the underlying basis for the overall essay structure. That is, there is an easily discernible logical flow from one idea (subtopic) to the next. Development of ideas is not interrupted by digressions of thought or extraneous material.

   Specifically:

   a. The plan for organizing the essay is clearly established in the beginning, e.g., through reference to the subtopics or line(s) of reasoning to be developed
   AND/OR
   The plan for organizing the essay is readily evident to the reader due to the effective use of paragraphs, transitions and linking expressions.

   b. The plan by which the essay is organized is consistently applied throughout the essay. All major subtopics (main points) clearly related to the main idea of the whole essay. There are no digressions of thought or extraneous material.

   c. The overall structure of the essay reflects some logical, underlying organizing rubric. That is, subtopics are presented according to some logical scheme, or order, such as: order of importance, seriousness; comparison and contrast; from specific to general; cause to effect; through classification.

   d. All statements related to the same idea (subtopics) are contiguous, i.e., are presented together as logically related units or blocks of thought.

3. The essay is organized according to some recognizable plan which is employed throughout the essay. Some organizing rubric is used to divide the content/ideas of the essay into units of thought. For example, each subtopic might represent a different reason or category of content. Subtopics, however, are not presented in any particular logical order.

   Specifically:

   a. The plan for organizing the essay is established in the beginning, e.g., through reference to the subtopic(s) or line(s) of reasoning to be developed
   AND/OR
   The plan for organizing the essay is evident to the reader due to the use of paragraphs, transitions and linking expressions.

(continued)

(continued)

b. The plan by which the essay is organized is employed throughout the essay. All subtopics relate to the one main idea. There are no major digressions; perhaps some minor digressions.

c. Major subtopics are not presented according to any underlying logical order; they could be "reshuffled" without dramatically altering the overall impact or clarity of the essay.

2. The essay is not structured according to any readily discernible plan. However, the reader is able to infer some association between major ideas. The logical flow from one idea to the next is so tenuous that the reader must supply his/her own links or transitions to form a clear relationship among ideas.

Specifically:

a. A plan for ordering ideas/details is attempted; the plan is not readily apparent to the reader. The essay does not make effective use of transitions or linking expressions to help the reader follow the trend of thought

AND/OR

b. The plan for ordering ideas/details is not consistently applied throughout the essay.

c. Some subtopics relate to the main idea. There may be major digressions of thought.

1. No plan for organizing ideas and details is apparent. Subtopics are not arranged in any discernible order. The reader has difficulty inferring any relationship among ideas or any pattern of thought.

Specifically:

a. No plan for organizing ideas and details is evident. Paragraphs, transitions and linking expressions, if present, are not helpful in following the writer's train of thought.

b. Few or no subtopics relate to the main idea and/or most or all of the essay is off the topic.

APPENDIX D

Expository Scale

ELEMENT 4

Support

Generaliztions and assertions are supported by specific, clear supporting statements.

4. All generalizations and assertions in this paper have logically related support. Supporting statements clearly are at a greater level of specificity than the generalizations they are intended to support. The writer makes effective use of concrete detail to support ideas.

   Specifically:

   a. Generalizations and assertions are supported by more specific statements.

   b. Supporting statements provide specific detail, such as illustrations, examples, facts, anecdotes, and/or employ concrete language to convey ideas.

3. Most generalizations and assertions are supported by more specific, clear and logically related statements.

   Specifically:

   a. Generalizations and assertions are supported by more specific statements.

   b. Many supporting statements provide specific detail, such as examples, facts, ancecdotes, and/or employ concrete language to convey ideas.

2. Support for many generalizations and assertions in this paper is not especially convincing. Supporting statements lack precision, either in language or in the use of specific detail.

   Specifically:

   a. An attempt is made to support generalizations and assertions through the use of statements at a greater level of specificity than the generalization they are intended to support.

   b. Supporting statements lack precision and clarity. The use of specific detail, such as examples, facts, or anecdotes, may be lacking and/or the language may be vague.

1. Support is provided for few or none of the generalizations or assertions. Few (or none) of the supporting statements are at a greater level of specificity than the generalizations themselves.

   Specifically:

   a. Supporting statements are provided for few or none of the generalizations or assertions.

   b. Supporting details, if present, are vague, confusing, or not logically related to the generalizations they apparently are intended to support, and/or the language of supporting statements is imprecise.

Expository Scale

ELEMENT 5

Mechanics

The essay is free of intrusive usage and mechanical errors.

4. The writer appears to have control over the usage and mechanical aspects of this essay.

   Specifically:

   a. There are only a few minor errors in usage and mechanics or no errors at all.

   b. Errors, if present, do not interfere with the clarity of communication.

3. Usage and mechanics are not a problem in this paper. However, the writer appears to be occasionally careless and makes a few minor errors which are readily evident to the discerning reader. The paper also may contain a few common errors in usage and mechanics.

   Specifically:

   a. There are only a few obvious usage or mechanical errors.

   b. Errors do not interfere with the clarity of communication, e.g., occasional confusion of subject-object pronouns ("between you and I..."), spelling errors on difficult words, misuse of colons and semicolons.

   c. There are no major errors, such as run-on sentences, inappropriate fragments, lack of subject-verb agreement in simple sentences.

2. This essay is flawed by errors in mechanics, although the paper does not strike the discerning reader as being illiterate.

   Specifically:

   a. There are obvious errors in mechanics throughout much of the essay.

   b. Errors occasionally detract from the clarity of the communication, such as confusing antecedents, consistent omission of key words.

   c. Many of the errors are major, e.g., many run-on sentences or inappropriate fragments, lack of subject-verb agreement in simple sentences.

1. Mechanical errors make this paper very difficult to read and understand.

   Specifically:

   a. There are excessive and obvious errors in mechanics throughout the essay; nearly every sentence contains some type of error.

   b. Errors consistently interfere with the writer's attempt to communicate.

   c. Errors are not restricted to one type of problem, such as run-on sentences. Errors are diverse in nature and major.

APPENDIX D

EXPOSITORY SCALE

Abbreviated Version

2. Essay Focus

In the introduction and/or conclusion:

4 =
- The subject is identified.
- The main idea is stated or implied.
- The topic is limited through reference to key points or lines of reasoning.

3 =
- The subject is identified.
- The main idea is stated or implied.
- The topic is partly limited by indicating number or type of key points.

2 =
- The subject is identified.
- The main idea is identified or implied.
- No attempt to limit topic, focus, give direction.

1 =
- The subject is not clear

  and/or
- The main idea is not stated or implied.
- Completely off the topic.

3. Essay Organization

4 =
- Plan clearly established in beginning through reference to subtopics or lines of reasoning.
- Effective use of paragraph transitions.
- Plan applied throughout the essay.
- Plan is based on a logical scheme; subtopics presented in a logical order, all related to main idea.

3 =
- Plan clearly established in beginning and/or clearly evident due to paragraphs, transitions.
- Plan applied throughout the essay.
- Subtopics not presented in any particular order.
- No major digressions, some minor.

2 =
- Plan is attempted, but not readily apparent.
- Plan can be inferred, but reader must supply lines of reasoning

  and/or
- Plan not followed throughout essay.
- Some subtopics relate to main idea; may be major digressions.

1 =
- No plan is attempted or apparent.
- Very few or not subtopics relate to main idea.

(continued)

(continued)

4. Support

4 = ˙Generalizations/assertions supported by specific statements.
˙Support is specific detail and/or uses concrete, precise language.

3 = ˙Most generalizations/assertions supported by specific statements.
˙Most support is specific detail and/or uses concrete, precise language.

2 = ˙An attempt is made to support generalizations/assertions.
˙Support lacks precision, clarity in details and/or language.

1 = ˙Support is provided for few or no generalizations/assertions.
˙Support, if present, is vague, confusing or not logically related to generalizations.

5. Usage and Mechanics

4 = ˙Few or no errors; errors not obvious.
˙Errors do not interfere with communication.

3 = ˙A few obvious errors.
˙Errors do not interfere with communication and are common minor ones.
˙No major errors.

2 = ˙Obvious errors throughout much of essay, though not illiterate.
˙Errors occasionally detract from clarity of communication.
˙Some major errors.

1 = ˙Excessive and obvious errors throughout essay.
˙Errors consistently interfere with communication.
˙Errors are diverse and major.

APPENDIX D

Factual Narrative Scale

ELEMENT 1

Impressionistic Rating Procedures

The purpose of Impressionistic Rating is to form a single impression of a piece of writing as to how well it communicates a whole message to the reader. Impressionistic scoring assumes that each characteristic that makes up an essay -- organization of ideas, content, mechanics and so on -- is related to all other characteristics. Impressionistic scoring further assumes that some qualities of an essay cannot easily be separated from each other. In short, the procedure views a piece of writing as a total work, the whole of which is greater than the sum of its parts.

Discerning readers naturally will attend to, or be influenced by, some essay characteristics more than others. In the Impressionistic scoring, however, readers should arrive at a judgment regarding the overall quality of the essay.

For this element, you are being asked to form an overall impression concerning the effectiveness of the essays as examples of narrative writing.

The Topics

You will be reading essays on two different topics. Both topics were designed to elicit writing in the narrative mode. Some views on narration are given below:

- Narration is the kind of discourse that tells what happened, how it happened.
- Narration gives an impression of movement in time.
- Narrative action includes time, logic, and meaning.

General Impression

Read each essay quickly in order to form an overall impression of its quality. To assign the essay a score, consider the following question: To what extent is the essay an example of effective narration?

Assign each paper a mark of 1 - 4 using the scale below.

4 = An excellent example of narration

3 = A good or adequate example of narration

2 = A minimally adequate example of narration

1 = A poor example of narration, barely readable and/or off the topic
0 = Is not a narrative

APPENDIX D

Factual Narrative Scale

ELEMENT 2

Essay Focus

The beginning or end of the essay clearly identifies the subject and main point of the whole essay.

4. The beginning (and/or conclusion) of this paper clearly implies or states the main points of the whole essay. It also limits the topic by alerting the reader to the key events or stages covered in the body of the essay.

   Specifically, at the beginning (and/or end):

   a. The subject of the essay is clearly identified.

   b. The main point of the whole essay is clearly stated or implied.

   c. The topic is clearly limited. That is, key points (e.g., major events) or major line(s) of action treated in the essay are identified, summarized or implied.

3. The beginning (and/or end) of this paper implies or states the main point of the whole essay. It sets limits on the topic, but does not clearly suggest how the main point is developed.

   Specifically, in the beginning (and/or end):

   a. The subject of the essay is clearly identified.

   b. The main point of the whole essay is clearly stated or implied.

   c. An attempt is made to limit the topic. That is, the number -- or general type -- of key points is indicated, but there is not clear reference to the important stages, or events treated in the body of the essay.

2. The beginning (and/or end) of this paper gives the reader a fairly clear sense of the main point of the whole essay. However, neither the introduction nor the conclusion help focus -- or bring direction to -- the body of the paper.

   Specifically, in the beginning (and/or end):

   a. The subject of the essay is identified.

   b. The main point of the whole essay is stated or implied.

   c. No attempt is made to limit the topic.

(continued)

ELEMENT 2 (continued)

1. Neither the beginning nor the end is helpful to the reader in obtaining any sense of the main point of the essay.

   Specifically, in the beginning (and/or end):

   a. The subject of the essay is not clearly identified or there is no reference to the subject

      AND/OR

   b. The main point of the whole essay is not clearly stated or implied or no reference is made to the main point or the reference is confusing.

APPENDIX D

Factual Narrative Scale

ELEMENT 3

Organization

4. The temporal (chronological) order of events is clear.

    a. Transitions and linking expressions enhance a sense of movement. Events and ideas are ordered according to a readily discernible logical progression.

    b. All events and ideas clearly relate to the main point of the action.

    c. There are no digressions or extraneous material.

3. The temporal order of events is clear.

    a. Transitions and linking help establish a sense of movement.

    b. Events and ideas are ordered according to a discernible logical progression.

    c. Most or all events and ideas relate to the main point of the essay, there are minor digressions or extraneous material.

2. The temporal order of events is not consistently clear although the essay does have a discernible beginning and end.

    a. Events and ideas are not ordered according to a discernible logical progression or sequence. Events may be merely listed.

    b. The relationship of many events and ideas to the main point is not consistently evident. There are major digressions and/or extraneous material.

1. There is no clear temporal order.

    a. Transitions are not present or are ineffective.

    b. Events and ideas are in a confusing order. The reader has a difficult time inferring the relationship among events and ideas. The paper may be expository or descriptive.

APPENDIX D

Factual Narrative Scale

ELEMENT 4

Support

Generalizations and assertions are supported by specific, clear supporting statements.

4. All general points and assertions in this paper have logically related details. Supporting statements clearly are at a greater level of specificity than the generalizations they are intended to support. The writer makes effective use of concrete detail to support ideas.

   Specifically:

   a. Generalizations and assertions are supported by more specific statements.

   b. Supporting statements provide specific detail, such as examples, facts, anecdotes, and/or employ concrete language to convey ideas about events, actions, and/or characters. A character is described by reference to appearance, feelings, thoughts and actions.

3. Most generalization points and assertions are supported by more specific, clear and logically related details.

   Specifically:

   a. General points and assertions are supported by more specific statements.

   b. Many supporting statements provide specific detail, such as examples, facts, anecdotes, and/or employ concrete language to convey ideas about events, actions and/or character(s)' appearance, feelings, thoughts and actions.

2. Support for many general points and assertions in this paper is not especially convincing. Supporting statements lack precision, either in language or in the use of specific detail.

   Specifically:

   a. An attempt is made to support general points and assertions through the use of statements at a greater level of specificity than the generalization they are intended to support.

   b. Supporting statements lack precision and clarity. The use of specific detail, such as examples, facts, anecdotes, character description may be lacking and/or the language may be vague.

(continued)

ELEMENT 4 (continued)

1. Support is provided for few or none of the general points or assertions. Few (or none) of the supporting statements are at a greater level of specificity than the generalizations themselves.

   Specificially:

   a. Supporting statements are provided for few or none of the general points or assertions.

   b. Supporting details, if present, are vague, confusing, or not logically related to the generalizations they apparently are intended to support, and/or the language of supporting statements is imprecise. There is little or no attempt to delineate a character's appearance, thoughts, feelings or actions.

Factual Narrative Scale

ELEMENT 5

Mechanics

The essay is free of intrusive usage and mechanical errors.

4. The writer appears to have control over the usage and mechanical aspects of this essay.

   Specifically:

   a. There are only a few minor errors in usage and mechanics or no errors at all.

   b. Errors, if present, do not interfere with the clarity of communication.

3. Usage and mechanics are not a problem in this paper. However, the writer appears to be occasionally careless and makes a few minor errors which are readily evident to the discerning reader. The paper also may contain a few common errors in usage and mechanics.

   Specifically:

   a. There are only a few obvious usage or mechanical errors.

   b. Errors do not interfere with the clarity of communication, e.g., occasional confusion of subject-object pronouns ("between you and I..."), spelling errors on difficult words, misuse of colons and semicolons.

   c. There are no major errors, such as run-on sentences, inappropriate fragments, lack of subject-verb agreement in simple sentences.

2. This essay is flawed by errors in mechanics, although the paper does not strike the discerning reader as being illiterate.

   Specifically:

   a. There are obvious errors in mechanics throughout much of the essay.

   b. Errors occasionally detract from the clarity of the communication, such as confusing antecedents, consistent omission of key words.

   c. Many of the errors are major, e.g., many run-on sentences or inappropriate fragments, lack of subject-verb agreement in simple sentences.

1. Mechanical errors make this paper very difficult to read and understand.

   Specifically:

   a. There are excessive and obvious errors in mechanics throughout the essay; nearly every sentence contains some type of error.

   b. Errors consistently interfere with the writer's attempt to communicate.

   c. Errors are not restricted to one type of problem, such as run-on sentences. Errors are diverse in nature and major.

Factual Narrative Scale

Abbreviated Version

2. Essay Focus

4 = ˙Subject is clearly identified.
˙Main point is stated or implied.
˙Topic limited, key points, events summarized or implied.

3 = ˙Subject clearly identified.
˙Main point stated or implied.
˙Attempt to limit topic by reference to number or general type of key points, but no reference to particular key events, stages.

2 = ˙Subject identified.
˙Main point stated or implied.
˙No attempt to limit topic, bring focus, direction.

1 = ˙Subject unclear, not identified.
˙Main point not stated or implied.
˙Reference to main point very confusing.

3. Organization

4 = ˙Temporal order clear, sense of movement.
˙Events and ideas have logical progression.
˙All events, points relate to main point.
˙No digression.

3 = ˙Temporal order clear, sense of movement.
˙Logical progression of events, ideas.
˙Most or all events and ideas relate to main point, minor digressions.

2 = ˙Temporal order not consistently clear, but has beginning, end.
˙Not logical progression of events, ideas. May be listed.
˙Not all events, ideas related to main point.
˙Major digressions or extraneous material.

1 = ˙No clear temporal order.
˙Confusing order, relationship among ideas, events.
˙Many digressions, extraneous material.
˙May be expository or descriptive.

Factual Narrative Scale

Abbreviated Version

4. Support

4 = ˙All general points have logically related details.
˙Support at greater level of specification.
˙Details are concrete examples, anecdotes, character description.

3 = ˙Most generalizations supported.
˙Many supporting statements, specific detail.

2 = ˙Unconvincing support for many general points.
˙Support lacks precision, concrete detail.

1 = ˙Little or not support.
˙Support not more specific than general points, events.
˙Support vague, not logically related.

5. Usage and Mechanics

4 = ˙Few or no errors; errors not obvious.
˙Errors do not interfere with communication.

3 = ˙A few obvious errors.
˙Errors do not interfere with communication and are common minor ones.
˙No major errors.

2 = ˙Obvious errors throughout much of essay, though not illiterate.
˙Errors occasionally detract from clarity of communication.
˙Some major errors.

1 = ˙Excessive and obvious errors throughout essay.
˙Errors consistently interfere with communication throughout most.
˙Errors are diverse and major.

APPENDIX E

DOMAIN SPECIFICATIONS FOR OBJECTIVE TEST

I. Expository Items

A. Main Idea/Focus

1. Objective: given a 100-word passage with a stated main idea, the student will select the main idea.

2. Stimulus: Expository passage. A 3-4 sentence paragraph of approximately 100-150 words will be written to explain a topic of interest and appropriate to senior high school students. The passage will include one sentence stating a generalization (main idea) which may appear as one of the first or last two sentences. Supporting sentences should be more specific than the main idea, presenting subparts, relevant details or illustrations. Passage structures will be primarily enumeration, comparison/contrast or examples.

3. Response Options.
   a. Correct Response: The correct response will be a sentence which:

      1) is a generalization or assertion superordinate to the other sentences in the passage.

      2) specifies the topic and main point/perspective made about the topic in the paragraph.

      3) is stated verbatim in the first or last two sentences of the passage

   b. Distractors: The distractors will be sentences with the following characteristics:

      1) a generalization relating to only one of the supporting sentences, for example: "Sea others are also an endangered species" - "There aren't many sea creatures which may become extinct."

      2) one of the supporting sentences in the paragraph.

      3) a new sentence presenting another relevant supporting statement. e.g.: "The bald eagle has long been protected."

      4) a new sentence tangential or irrelevant to the topic of the passage. e.g., "What a difference!" "I think sea otters are cute."

B. Supporting Detail

1. Objective: Students will select a new sentence which would also support the passage main idea.

2. Stimulus: Same passage as described for main idea question.

3. Response Options:

    a. Correct Response: The correct sentence will be one which is:

        1. a new sentence derived from the stated main idea of the passage

        2. maintains the same relationship to other supporting sentences (e.g., is another example or another contrast)

    b. Distractors: The distractors will be:

        1. A new sentence making a statement more general than any of the supporting sentences, parallel in generality to the main idea sentence.

        2. A statement irrelevant to the main idea.

        3. A statement related to only one of the other supporting details.

C. Organization

1. Objective: The student will select a new sentence which would best fit in the point of the passage indicated by an arrow.

2. Stimulus: Same expository passage.

3. Response Questions:

    a. Correct Response: The correct sentence will be a new sentence which :

        1. Adds to, elaborates upon the immediately preceding sentence.

        2. Makes sense in relation to the following sentence.

        3. Is at the same level of specificity as the other supporting sentences.

b. Distractors: The distractors will:

1. Be one sentence presenting a new generalization.
2. A detail which does not follow from - relate to the preceding sentence, but would fit elsewhere in the passage.
3. A detail irrelevant or tangential to the main idea.

D. Sample Item(s)

The sample items present the format, directions, and illustrative items for each of the three objectives.

In recent years learning to ski has become much easier. The traditional method was to put beginners on very long skis and to teach them to ski in three different methods [1]. Each stage in the traditional approach required "unlearning" techniques from the earlier stage. Luckily modern ski schools allow beginners to progress much more quickly. Students use very short skis which are easier to turn and stop. Within two or three lessons beginners practice using "advanced" methods to turn and stop. What a difference!

1. Which sentence best states the main idea of the passage?

____a. Students use short skis which are easier to turn and stop.

____b. Each new method or stage required unlearning techniques from the earlier stage.

____c. What a difference!

____d. In recent years, learning to ski has become much easier.

2. Which of these new details would also best support the main idea?

____a. Skiing is good exercise as well as being fun.

____b. The newly designed equipment has helped many people learn to ski in only one or two lessons.

____c. Experts prefer longer skis for speed and stability.

____d. Running and climbing stairs will help you get in shape for skiing.

3. Which of these new sentences would fit best if added where the arrow [↓] appears?

   ____a. The traditional stages were to learn the "snow plow" turn, then a stem christie turn and finally to keep the skis parallel when turning.

   ____b. Children as young as two and three can learn to ski.

   ____c. Expert skiiers just speed straight down the hill.

   ____d. One new approach gradually increases the length of skis beginners use.

II. Narrative Items

A. Main Idea/Focus

1. Objective: Given a 100-word passage with a stated main idea, the student will select that main idea sentence.

2. Stimulus: Narrative passage. A 3-4 sentence paragraph of approximately 100-150 words be written to portray how an event or stages in a process occurred. The passage will include a generalization summarizing the point of the event and 2-3 sentences describing separate events, details in the course of events. All passages will use chronological, temporal ordering. No foreshadowing or flashback will be included.

3. Response Options

   a. Correct Response: The correct main idea sentence will be one summarizing the point of the narrative and have been stated in the first or last two sentences of the passage.

   b. Distractors: The incorrect sentence will be:

      1. A generalization not summarizing all of the supporting details.

      2. A supporting sentence from the passage.

      3. A new supporting detail related to the passage.

B. Supporting Detail

1. Objective: The student will select a new sentence which would also support the passage main idea.

2. Stimulus: Same narrative passage.

3. Response Options

   1. Correct Response: The correct sentence will be at the same level of specificity as the other supporting sentences, fit into the temporal order, and relate to the topic.

   2. Distractors: The distractor will include:

      a. A new statement related to the topic but more general than other supporting details.

      b. A new detail irrelevant to the passage

      c. A new detail relating to only one of the other details, not to the passage focus.

C. Organization

1. Objective: The student will select the new sentence which would best fit at the point in the passage indicated by the arrow.

2. Stimulus: Same narrative passage.

3. Response Options

   a. Correct Response: The correct sentence will be at the same level of specificity as the other supporting sentences, relevant to the topic and will fit logically and temporally between the sentences indicated by the arrow.

D. Sample Items

The sample items present the format, directions, and illustrative items for each of the three objectives related to a narrative passage.

Many years ago only Indians and Eskimos lived in our most northern state, Alaska. Then, people from other countries began to go to the northern land. [↓] Some people wanted to explore the vast country. Others went to hunt for seals and polar bears. Then, about seventy-five years ago, gold was discovered in Alaska, and thousands rushed to search for the valuable metal. The population of Alaska now consists of many different peoples.

1. Which sentence best states the main idea of the passage?

    ____a. Others went to hunt for seals and polar bears.

    ____b. The population of Alaska now consists of many different peoples.

    ____c. Some people wanted to explore the vast country.

    ____d. Many years ago only Indians and Eskimos lived in our most northern state, Alaska.

2. Which of these new details would also best support the main idea?

    ____a. Alaska and Hawaii were the last territories to become states.

    ____b. Eskimo art is increasingly recognized as very valuable.

    ____c. Alaska's fight for statehood is an interesting story.

    ____d. Alaska's liquid gold, oil, also drew new population.

3. Which of these new sentences would fit best if added where the arrow [↓] appears?

    ____a. Eskimo art is increasingly recognized as very valuable.

    ____b. Alaska's fight for statehood is an interesting story.

    ____c. They came for many reasons.

    ____d. Alaska's liquid gold, oil, has made the state very rich.

APPENDIX E

Objective Test

YOUR NAME: ______________________ YOUR SCHOOL ______________________

YOUR ENGLISH TEACHER: ______________________ DATE: ______________________

YOUR GRADE: ( ) 10th ( ) 11th ( ) 12th

Directions:

1. Please try to answer each question on the following pages.
2. Mark your answers in this booklet in pen or pencil.
3. Each question can be answered with a checkmark (✓) in the spaces provided.
4. Read the directions to each part very carefully.
5. Raise your hand for assistance if you do not understand what you are supposed to do.

Directions: Read each of the passages below and answer the questions under them. Place a check (__✓__) by only the one sentence you think is best.

Linda Ronstadt, whose songs have a strong beat, has become quite famous. Her fame may be due in part to the fact that many people enjoy songs with a clear melody. Popular female singers are now performing a variety of types of music. Some of those groups, like Heart, have musical styles which range from soft love songs to rhythm and blues [↓]. Some of the most frequently played disco music is performed by women such as Thelma Houston.

1. Which sentence best states the main idea of the passage?

____a. Teenagers usually prefer music with a driving rhythm.

____b. Female country music singers are doing well on the charts.

____c. Popular female singers are now performing a variety of types of music.

____d. Some of the most popular music has a good rhythm section.

2. Which of these new details would also best support the main idea?

____a. Rod Stewart has become quite popular too.

____b. Some female singers have had tragic lives.

____c. Helen Reddy has had a number of hit albums.

____d. Disco music is really my favorite.

3. Which of these new sentences would fit best where the arrow [↓] appears?

____a. Oliva Newton John's music also has a variety of styles.

___ b. Many new male singers are doing well on the charts.

____c. Women are becoming stars in sports, too.

____d. Linda Rondstat is still my favorite.

In recent years learning to ski has become much easier. The traditional method was to put beginners on very long skis and to teach them to ski in three different methods [↓]. Each stage in the traditional approach required "unlearning" techniques from the earlier stage. Luckily modern ski schools allow beginners to progress much more quickly. Students use very short skis which are easier to turn and stop. Within two or three lessons beginners practice using "advanced" methods to turn and stop. What a difference!

4. Which sentence best states the main idea of the passage?

____a. Students use short skis which are easier to turn and stop.

____b. Each new method or stage required unlearning techniques from the earlier stage.

____c. What a difference!

____d. In recent years, learning to ski has become much easier.

5. Which of these new details would also best support the main idea?

____a. Skiing is good exercise as well as being fun.

____b. The newly designed equipment has helped many people learn to ski in only one or two lessons.

____c. Experts prefer longer skis for speed and stability.

____d. Running and climbing stairs will help you get in shape for skiing.

6. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. The traditional stages were to learn the "snow plow" turn, then a stem christie turn and finally to keep the skis parallel when turning.

____b. Children as young as two and three can learn to ski.

____c. Expert skiiers just speed straight down the hill.

____d. One new approach gradually increases the length of skis beginners use.

Real gold never rusts. Regarded as a precious metal, gold has been highly valued throughout history. It is considered a beautiful metal because it is soft and can be moded into decorative shapes. The value of gold is another reason why people like it. [↓] Gold is valuable because there is not much of it. People tend to enjoy owning things which are considered rare.

7. Which sentence best states the main idea of the passage?

____a. Gold never rusts.

____b. Regarded as a precious metal, gold has been highly valued throughout history.

____c. The value of gold is another reason why people like it.

____d. It is considered a beautiful metal because it is soft and can be moded into decorative shapes.

8. Which of these new details would also best support the main idea?

____a. Diamonds are precious gems.

____b. Gold is mined by many different methods.

____c. Clearly the beauty of gold is one reason why people like it.

____d. Many miners have been disappointed to find only "fool's gold."

9. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. Beautiful jewlery is made from gold.

____b. Emeralds, diamonds and other precious stones are also rare.

____c. Gold is mined by many different methods.

____d. The price of gold has increased three-fold in the last decade.

Worker bees leave the hive early and do not return until the sun sets. They fly around most of the day, covering an average of ten miles per day. [↓] Bees called drones, however, lead an easy life. They stay in the hive, eat, and fertilize eggs. The queen bee lays all the eggs which will become new worker bees and drones. "Busy as a bee" is a saying which applies to only some types of bees in a hive.

10. Which sentence best states the main idea of the passage?

____a. They fly around most of the day covering an average of ten miles per day.

____b. "Busy as a bee" is a saying applying to only some types of bees in a hive.

____c. Bees called drones, however, lead an easy life.

____d. The queen bee lays all the eggs which will become new worker bees and drones.

11. Which of these new details would also best support the main idea?

____a. Drones are not very productive.

____b. Flowering shrubs benefit greatly from bees' activity.

____c. Bees can kill people who are allergic to their sting.

____d. The honey bees make is flavored by the type of flower they get the pollen from.

12. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. To make honey, worker bees spend hours each day chewing pollen.

____b. The queen bee is the largest bee in the hive.

____c. Ants, of course, can't make honey.

____d. Often people who are lazy are called drones.

Working at a fast food restaurant means working under a lot of pressure. Most customers want their food right away and get angry if they have to wait at all. [↓] The kitchen is usually noisy making it hard to hear the orders and get the food served quickly. Many things need to be done fast -- turn the hamburgers, pour the milk shakes, put the fries in a bag.

13. Which sentence best states the main idea of the passage?

____a. Most customers want their food right away and get angry if they have to wait at all.

____b. Many things need to be done fast -- turn the hamburgers, pour the milk shakes, put the fries in a bag.

____c. Working at a fast food restaurant means working under a lot of pressure.

____d. The kitchen is usually noisy making it hard to hear the orders and get the food served quickly.

14. Which of these new details would also best support the main idea?

____a. People who work in demanding jobs usually are not paid enough.

____b. Often fast food workers have to ring the cash register and serve food at the same time.

____c. Fast food chains specialize in different kinds of foods.

____d. The bags of fries sometimes drip grease on my clothes.

15. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. Each cup has to have a lid securely in place.

____b. Customers waiting in their cars at drive-through restaurants sometimes honk their horns or shout.

____c. People who work in demanding jobs usually are not paid enough.

____d. Fast food chains specialize in different types of food.

The African Gypsy moth, like most insects, goes through several stages before it is a full grown moth. The first stage is an egg. When the egg hatches the larva stage begins. The larva looks like a caterpillar and eats great numbers of leaves. It then goes into the pupa stage where it spins a cocoon. [↓] After a number of weeks the pupa cracks open and the Gypsy moth crawls out.

16. Which sentence best states the main idea of the passage?

____a. After a number of weeks the pupa cracks open and the Gypsy moth crawls out.

____b. Its large red and yellow wings make the African Gypsy moth one of the most beautiful insects.

____c. It then goes into the pupa stage where it spins a cocoon.

____d. The African Gypsy moth, like most insects, goes through several stages before it is a full grown moth.

17. Which of these new details would also best support the main idea?

____a. In the third pupa stage, the larva again changes form.

____b. Insects develop into beautiful and ugly creatures.

____c. Moths once ate holes in my wool sweater.

____d. The Gypsy moth has often been described as a "flying sunset."

18. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. The Gypsy moth has often been described as a "flying sunset."

____b. The cocoon looks like a large lifeless peapod.

____c. Within a very short time this moth may lay eggs and the process will begin again.

____d. The cocoons of insects assume many interesting forms.

Many years ago only Indians and Eskimos lived in our most northern state, Alaska. Then, people from other countries began to go to the northern land. [↓] Some people wanted to explore the vast country. Others went to hunt for seals and polar bears. Then, about seventy-five years ago, gold was discovered in Alaska, and thousands rushed to search for the valuable metal. The population of Alaska now consists of many different peoples.

19. Which sentence best states the main idea of the passage?

____a. Others went to hunt for seals and polar bears.

____b. The population of Alaska now consists of many different peoples.

____c. Some people wanted to explore the vast country.

____d. Many years ago only Indians and Eskimos lived in our most northern state, Alaska.

20. Which of these new details would also best support the main idea?

____a. Alaska and Hawaii were the last territories to become states.

____b. Eskimo art is increasingly recognized as very valuable.

____c. Alaska's fight for statehood is an interesting story.

____d. Alaska's liquid gold, oil, also drew new population.

21. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. Eskimo art is increasingly recognized as very valuable.

____b. Alaska's fight for statehood is an interesting story.

____c. They came for many reasons.

____d. Alaska's liquid gold, oil, has made the state very rich.

The Sahara Desert is a very large desert in northern Africa. The discovery of oil has brought great change to the people who live in the Sahara. Before the oil men came, the desert tribes wandered from oasis to oasis. [↓] Now high-paying jobs in the oil fields have encouraged tribes people to settle into permanent company-built towns. With water pumped into the towns the tribes can grow the food they once had to seek.

22. Which sentence best states the main idea of the passage?

    ____a. Now high-paying jobs in the oil fields have encouraged tribes people to settle into permanent company-built towns.

    ____b. With water pumped into the towns the tribes can grow the food for which they once had to wander and search.

    ____c. The discovery of oil has brought great change to the people who live in the Sahara.

    ____d. The Sahara Desert is a very large desert in northern Africa.

23. Which of these new details would also best support the main idea?

    ____a. Whole tribes have sold their camels and abandoned their traditions.

    ____b. Nomadic tribes of the earth have almost disappeared.

    ____c. The search for liquid gold lures men to harsh exotic lands.

    ____d. Many Americans and Europeans have come to the oil fields.

24. Which of these new sentences would fit best if added where the arrow [↓] appears?

    ____a. Each town boasted rows and rows of neat, square little houses.

    ____b. Exposure to western ways threatened customs of the tribes.

    ____c. By trading goods carried on their camels, the tribes could get fresh water and food.

    ____d. Of the world's deserts, the Sahara may be the hottest and harshest.

When Emily heard the doctor's report, she was stunned. She might lose the sight in her right eye. In reaction to her personal crisis, Emily's behavior went through a series of dramatic changes. For weeks she withdrew from her family and friends, brooding, refusing to admit her terror. Then she became supercharged, hurtling from activity to activity, pretending nothing was wrong. [↓] At last Emily confided in her friend, releasing her torrent of fear. From that point Emily seemed to regain her balance and slowly face her fear.

25. Which sentence best states the main idea of the passage?

____a. In reaction to her personal crisis, Emily's behavior went through a series of dramatic changes.

____b. At last Emily confided in her friend, releasing her torrent of fear.

____c. For weeks she withdrew from her family and friends brooding, refusing to admit her terror.

____d. When Emily heard the doctor's report she was stunned.

26. Which of these new details would also best support the main idea?

____a. Emily had always loved to paint.

____b. A number of serious eye conditions, such as glaucoma, can threaten eyesight.

____c. A family's reactions to personal crisis can vary in many ways.

____d. Emily's usually high grades dropped drastically.

27. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. Family ties are disintegrating in modern society.

____b. First Emily couldn't believe it.

____c. The busier she kept, the easier it was to forget.

____d. Emily had always loved to paint.

Scuba diving lessons usually begin in a swimming pool. First the instructor helps students calculate air pressure readings and demonstrates how to clear the face mask and use the air tanks. Then beginners practice adjusting the equipment and using safety procedures. At last the class goes on a series of ocean dives in progressively deeper water. [↓] Scuba diving instruction is designed to guide the beginners' development safely.

28. Which sentence best states the main idea of the passage?

____a. Then beginners practice adjusting the equipment and using safety procedures.

____b. Scuba diving instruction is designed to guide the beginners' development safely.

____c. At last the class goes on a series of ocean dives in progressively deeper water.

____d. Beginning lessons in scuba diving are usually in a swimming pool.

29. Which of these new details would also best support the main idea?

____a. Some plants have dangling and waving leaves in which the diver can get caught.

____b. If a diver surfaces too quickly he may get a painful condition called the "bends."

____c. The small area and shallowness of the pool allow the instructor to watch beginners carefully.

____d. Scuba diving equipment has improved greatly in recent years.

30. Which of these new sentences would fit best if added where the arrow [↓] appears?

____a. Students love the opportunities to try their new skills and experience the beauty of underwater creatures.

____b. Most fish are gentle but may occasionally try to attack a diver.

____c. Scuba diving equipment has improved greatly in recent years.

____d. Some plants have dangling and waving leaves in which a diver can get caught.

APPENDIX F

Analysis of Variance Table for Alpha Coefficient: Pilot Test
(Raters 1, 2, 3, 4)

Expository Scale: General Impression

Source	SS	DF	F	P
Between S's	31.659	19		
Within S's	9.188	60		
Between Measures (raters)	1.609	3	4.03	.01
Residual	7.578	57		

Alpha = .92 Standardized Item (S.I.) alpha = .93

Expository Scale: Focus

Between S's	27.609	19		
Within S's	15.563	60		
Between Measures (raters)	6.784	3	12.52	.000
Residual	9.378	57		

Alpha = .89 (S.I.) alpha = .89

Expository Scale: Organization

Between S's	30.309	19		
Within S's	12.563	60		
Between Measures (raters)	3.109	3	6.24	.00
Residual	9.453	57		

Alpha = .90 (S.I.) alpha = .89

Expository Scale: Support

Between S's	25.050	19		
Within S's	11.250	60		
Between Measures (raters)	3.325	3	7.971	.00
Residual	7.925	57		

Alpha = .89 (S.I.) alpha = .90

Expository Scale: Mechanics

Between S's	33.559	19		
Within S's	20.188	60		
Between Measures (raters)	6.309	3	8.63	.00
Residual	13.878	57		

Alpha = .86 (S.I.) alpha = .84

APPENDIX F

Analysis of Variance Table for Alpha Coefficient: Pilot Test (Raters 1, 2, 3, 4)

Narrative Scale: General Impression

Source	SS	DF	F	P
Between S's	49.557	11		
Within S's	10.187	36		
Between Measures (raters)	1.516	3	1.92	.145
Residual				

Alpha = .94 Standardized Item (S.I.) alpha = .95

(SPSS Batch Release 7.8 update manual, March, 1977
Program: Reliability)

Narrative Scale: Focus

Source	SS	DF	F	P
Between S's	18.932	11		
Within S's	9.438	36		
Between Measures (raters)	1.770	3	2.404	.085
Residual	8.104	33		

Alpha = .87 (S.I.) alpha = .88

Narrative Scale: Organization

Source	SS	DF	F	P
Between S's	24.938	11		
Within S's	9.875	36		
Between Measures (raters)	1.770	3	2.404	.085
Residual	8.104	33		

Alpha = .89 (S.I.) alpha = .93

Narrative Scale: Support

Source	SS	DF	F	P
Between S's	17.729	11		
Within S's	6.250	36		
Between Measures (raters)	.854	3	1.74	.177
Residual	5.395	33		

Alpha = .90 (S.I.) alpha = .92

Source	SS	DF	F	P
Between S's	9.682	11		
Within S's	16.437	36		
Between Measures (raters)	8.516	3	11.82	.0002
Residual	7.922	33		

Alpha = .73 (S.I.) alpha = .92

APPENDIX F

Data for Estimating Generalizability Coefficients: Pilot Test

Expository Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.327	19	3.333	
Topics	.034	1	3.306	
Raters	.026	3	1.107	.74
Subjects x Topics	.111	19	.806	
Subjects x Raters	.047	57	.266	
Topics x Raters	.012	3	.106	
Subjects x Topics x Raters	.360	57	.360	

Expository Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.241	19	2.906	
Topics	.011	1	.156	
Raters	.094	3	4.123	.63
Subjects x Topics	.162	19	1.090	
Subjects x Raters	.057	57	.329	
Topics x Raters	.001	3	.456	
Subjects x Topics x Raters	.443	57	.442	

Expository Scale: Organization

Source	$E\sigma^2$	DF	MS	G
Subjects	.297	19	3.190	
Topics	.009	1	.625	
Raters	.045	3	2.072	.74
Subjects x Topics	.121	19	.756	
Subjects x Raters	.070	57	.332	
Topics x Raters	.002	3	.223	
Subjects x Topics x Raters	.271	57	.271	

APPENDIX F

Expository Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.241	19	2.637	
Topics	.055	1	4.900	
Raters	.054	3	2.217	
Subjects x Topics	.107	19	.716	.77
Subjects x Raters	.004	57	.278	
Topics x Raters	.011	3	.500	
Subjects x Topics x Raters	.287	57	.287	

Expository Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.323	19	3.532	
Topics	.009	1	.756	
Raters	.098	3	4.206	
Subjects x Topics	.114	19	.782	.73
Subjects x Raters	.081	57	.486	
Topics x Raters	.010	3	.123	
Subjects x Topics x Raters	.325	57	.325	

APPENDIX F

Data for Estimating Generalizability Coefficients: Pilot Test

Narrative Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.889	11	9.010	
Topics	.030	1	.104	
Raters	.038	3	1.010	.86
Subjects x Topics	.342	11	1.874	
Subjects x Raters	.010	33	.525	
Topics x Raters	.036	3	.660	
Subjects x Topics x Raters	.505	33	.505	

Narrative Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.333	11	3.442	
Topics	.018	1	.104	
Raters	.039	3	1.539	.76
Subjects x Topics	.087	11	.692	
Subjects x Raters	.044	33	.432	
Topics x Raters	.013	3	.510	
Subjects x Topics x Raters	.344	33	.344	

Narrative Scale: Organization

Source	$E\sigma^2$	DF	MS	G
Subjects	.458	11	4.534	
Topics	.012	1	.417	
Raters	.032	3	1.180	
Subjects x Topics	.095	11	.700	.79
Subjects x Raters	.086	33	.491	
Topics x Raters	.007	3	.236	
Subjects x Topics x Raters	.319	33	.319	

Narrative Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.253	11	3.223	
Topics	.016	1	.166	
Raters	.019	3	.569	
Subjects x Topics	.218	11	1.166	.76
Subjects x Raters	.016	33	.327	
Topics x Raters	.018	3	.833	
Subjects x Topics x Raters	.295	33	.295	

Narrative Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.131	11	1.760	
Topics	.003	1	.937	
Raters	.227	3	5.677	
Subjects x Topics	.059	11	.503	.52
Subjects x Raters	.106	33	.480	
Topics x Raters	.021	3	.104	
Subjects x Topics x Raters	.268	33	.268	

APPENDIX G

Data for Estimating Generalizability Coefficients: Final Study

Raters 1 & 2

Expository Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.033	13	.815	
Topics	.010	1	3.02	
Raters	.021	1	.875	.22
Subjects x Topics	.115	13	.440	
Subjects x Raters	.120	13	.452	
Topics x Raters	-.014	1	-.018	
Subjects x Topics x Raters	.210	13	.210	

Expository Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.110	13	.946	
Topics	.099	1	3.02	
Raters	.005	1	.161	.60
Subjects x Topics	.115	13	.518	
Subjects x Raters	-.005	13	.276	
Topics x Raters	-.019	1	.018	
Subjects x Topics x Raters	.287	13	.287	

Expository Scale: Organization

Source	$E\sigma^2$	DF	MS	G
Subjects	.129	13	1.374	
Topics	.055	1	2.160	
Raters	.049	1	1.446	.61
Subjects x Topics	.302	13	.815	
Subjects x Raters	.022	13	.254	
Topics x Raters	-.014	-	-.018	
Subjects x Topics x Raters	.210			

(Continued)

Raters 1 & 2

Expository Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.159	13	1.153	
Topics	.000	1	.286	
Raters	-.549	1	.071	.68
Subjects x Topics	.107	13	.362	
Subjects x Raters	.007	13	.302	
Topics x Raters	-.005	1	.071	
Subjects x Topics x Raters	.148	13	.148	

Expository Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.164	13	1.133	
Topics	-.049	1	1.446	
Raters	-.003	1	.161	.57
Subjects x Topics	.022	13	.254	
Subjects x Raters	.110	13	.430	
Topics x Raters	-.014	1	.018	
Subjects x Topics x Rater	.210	13	.210	

APPENDIX G

Data for Estimating Generalizability Coefficients: Final Study

Raters 3 & 4

Expository Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.312	25	1.606	
Topics	-.002	1	.038	
Raters	.105	1	5.538	.83
Subjects x Topics	.060	25	.298	
Subjects x Raters	.029	25	.238	
Topics x Raters	-.007	1	0	
Subjects x Topics x Raters	.180	25	.180	

Expository Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.245	25	1.463	
Topics	-.01	1	.962	
Raters	.032	1	2.163	.70
Subjects x Topics	.029	25	.430	
Subjects x Raters	.026	25	.423	
Topics x Raters	.004	1	.471	
Subjects x Topics x Raters	.371	25	.371	

Expository Scale: Organization

Source	$E\sigma^2$	DF	MS	G
Subjects	.217	25	1.493	
Topics	-.004	1	.154	
Raters	.017	1	3.846	.67
Subjects x Topics	.100	25	.454	
Subjects x Raters	.086	25	.426	
Topics x Raters	-.004	1	.154	
Subjects x Topics x Raters	.253	25	.254	

(Continued)

Expository Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.538	25	.727	
Topics	.231	1	.471	
Raters	.039	1	2.163	.85
Subjects x Topics	.171	25	.591	
Subjects x Raters	.057	25	.363	
Topics x Raters	-.009	1	.961	
Subjects x Topics x Raters	.250	25	.250	

Expository Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.268	25	1.446	
Topics	-.002	1	0	
Raters	.166	1	8.653	.77
Subjects x Topics	.041	25	.400	
Subjects x Raters	-.012	25	.294	
Topics x Raters	-.011	1	.038	
Subjects x Topics x Rater	.318	25	.318	

APPENDIX G

Data for Estimating Generalizability Coefficients: Final Study

Raters 1 & 2

Narrative Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.457	20	3.840	
Topics	-.040	1	.476	
Raters	-.042	1	1.714	.86
Subjects x Topics	.874	20	2.048	
Subjects x Raters	-.018	20	.264	
Topics x Raters	-.014	1	0	
Subjects x Topics x Rater	.300	20	.30	

Narrative Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.083	20	.715	
Topics	.002	1	.107	
Raters	.046	1	2.011	.46
Subjects x Topics	-.002	20	.332	
Subjects x Raters	.025	20	.387	
Topics x Raters	-.015	1	.119	
Subjects x Topics x Raters	.337	20	.337	

Narrative Scale: Organization

Source	$E\sigma^2$	DF	MS	
Subjects	.262	20	1.608	
Topics	-.006	1	.048	
Raters	.008	1	.429	.65
Subjects x Topics	.053	20	.673	
Subjects x Raters	-.056	20	.454	
Topics x Raters	-.018	1	.190	
Subjects x Topics x Raters	-.565	20	.565	

APPENDIX G

Final Study

Narrative Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.283	20	1.511	
Topics	.595	1	.429	
Raters	.238	1	.476	.77
Subjects x Topics	.066	20	.479	
Subjects x Raters	-.050	20	.248	
Topics x Raters	-.014	1	.048	
Subjects x Topics x Rate	.348	20	.348	

Narrative Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.233	20	1.287	
Topics	.014	1	.583	
Raters	.001	1	.107	.71
Subjects x Topics	-.014	20	.308	
Subjects x Raters	.023	20	.382	
Topics x Raters	-.015	1	.119	
Subjects x Topics x Raters	.337	20	.337	

APPENDIX G

Data for Estimating Generalizability Coefficients: Final Study

Raters 3 & 4

Narrative Scale: General Impression

Source	$E\sigma^2$	DF	MS	G
Subjects	.877	17	5.104	
Topics	.361	1	14.222	
Raters	.607	1	.222	.88
Subjects x Topics	.583	17	1.663	
Subjects x Raters	-.034	17	.428	
Topics x Raters	-.025		.056	
Subjects x Topics x Rater	.497	17	.497	

Narrative Scale: Focus

Source	$E\sigma^2$	DF	MS	G
Subjects	.515	17	2.390	
Topics	-.015	1	.125	
Raters	-.003	1	.680	.85
Subjects x Topics	-.013	17	.213	
Subjects x Raters	.059	17	.357	
Topics x Raters	.024		.680	
Subjects x Topics x Raters	.239	17	.239	

Narrative Scale: Organization

Source	$E\sigma^2$	DF	MS	G
Subjects	.436	17	2.592	
Topics	.176	1	7.347	
Raters	.163	1	.347	.85
Subjects x Topics	.324	17	.906	
Subjects x Raters	-.029	17	.200	
Topics x Raters	.005	1	.347	
Subjects x Topics x Raters	.259	17	.259	

(Continued)

Narrative Scale: Support

Source	$E\sigma^2$	DF	MS	G
Subjects	.090	17	1.271	
Topics	.204	1	8.000	
Raters	.098	1	3.556	.53
Subjects x Topics	.256	17	.824	
Subjects x Raters	.044	17	.320	
Topics X Raters	-.001	1	.056	
Subjects x Topics x Rate	.232	17	.232	

Narrative Scale: Mechanics

Source	$E\sigma^2$	DF	MS	G
Subjects	.359	17	1.908	
Topics	-.005	1	.222	
Raters	.005	1	.500	.83
Subjects x Topics	.088	17	.369	
Subjects x Raters	.051	17	.294	
Topics x Raters	.002	1	.222	
Subjects x Topics x Raters	.193	17	.193	